GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: August 3, 2005, 10:02:38 ; Search time 167 Seconds
(without alignments)
963.427 Million cell updates/sec

Title: US-10-706-691-16
Perfect score: 416
Sequence: 1 MKRERGALSRASRALRLAPF.....TAGVHIIREQDEAGPVEISA 416

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 2105692 seqs, 386760381 residues

Word size : 0

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : A_Geneseq_16Dec04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	416	100.0	416	7	ABG75379	Abg75379 Predicted
2	416	100.0	416	7	ABG75377	Abg75377 Human INS
3	416	100.0	416	8	ADO47892	Ado47892 Human pro
4	416	100.0	416	8	ADS11056	Ads11056 Human the
5	383	92.1	383	8	ADO47895	Ado47895 Human mat
6	291	70.0	298	5	AAE14784	Aae14784 Human imm
7	268	64.4	270	8	ADO47887	Ado47887 Human pro
8	268	64.4	270	8	ADS11055	Ads11055 Human the
9	240	57.7	246	7	ABG75380	Abg75380 INSP052 e
10	235	56.5	237	8	ADO47890	Ado47890 Human mat
11	224	53.8	224	5	AAE26421	Aae26421 Human tra
12	214	51.4	256	8	ADM87341	Adm87341 Human pro
13	153	36.8	418	7	ABG75378	Abg75378 Murine IN
14	151	36.3	256	4	AAM24238	Aam24238 Human EST
15	151	36.3	256	8	ADM87787	Adm87787 Human EST
16	151	36.3	256	8	ADS12269	Ads12269 Human the
17	151	36.3	256	8	ADS12268	Ads12268 Human the
18	122	29.3	367	8	ADQ65357	Adq65357 Novel hum
19	114	27.4	114	7	ABG75371	Abg75371 Human INS
20	100	24.0	100	7	ABG75376	Abg75376 Human INS
21	94	22.6	188	7	ABG75372	Abg75372 Human INS
22	58	13.9	58	3	AAG01648	Aag01648 Human sec
23	33	7.9	33	8	ADO47889	Ado47889 Human sig
24	33	7.9	33	8	ADO47894	Ado47894 Human sig
25	31	7.5	31	7	ABG75373	Abg75373 Human INS
26	29	7.0	29	7	ABG75370	Abg75370 Human INS
27	25	6.0	25	7	ABG75374	Abg75374 Human INS
28	23	5.5	23	7	ABG75375	Abg75375 Human INS
29	9	2.2	63	7	ADJ83109	Adj83109 Immunoglo
30	9	2.2	63	7	ADJ83139	Adj83139 Immunoglo
31	9	2.2	63	8	ADK40849	Adk40849 Immunoglo
32	9	2.2	63	8	ADK40851	Adk40851 Immunoglo
33	9	2.2	333	7	ABO71364	Abo71364 Pseudomon
34	8	1.9	152	8	ADG22490	Adg22490 Cyanophag
35	8	1.9	179	7	ABO78158	Abo78158 Pseudomon
36	8	1.9	192	4	ABB62455	Abb62455 Drosophil
37	8	1.9	192	4	AAU38940	Aau38940 Drosophil
38	8	1.9	192	7	ADC35802	Adc35802 Drosophil
39	8	1.9	234	6	ABU31383	Abu31383 Protein e
40	8	1.9	303	6	ABU44830	Abu44830 Protein e
41	8	1.9	305	6	ABU47271	Abu47271 Protein e
42	8	1.9	305	6	ABU48104	Abu48104 Protein e
43	8	1.9	305	6	ABU15437	Abu15437 Protein e
44	8	1.9	338	7	ABO65915	Abo65915 Klebsiell
45	8	1.9	362	4	ABB62460	Abb62460 Drosophil

ALIGNMENTS

RESULT 1
ABG75379
ID ABG75379 standard; protein; 416 AA.
XX
AC ABG75379;
XX
DT 22-APR-2004 (first entry)
XX
DE Predicted INSP052 protein.
XX
KW INSP052; human; cell proliferation; autoimmune disease; inflammation;
KW cardiovascular disease; neurological disease; psychiatric disease;
KW developmental disease; metabolic disorder; infection;
KW immunoglobulin domain-containing cell surface recognition molecule.
XX
OS Unidentified.
XX
PN WO2003093316-A2.
XX
PD 13-NOV-2003.
XX
PF 30-APR-2003; 2003WO-GB001851.
XX
PR 30-APR-2002; 2002GB-00009884.
XX
PA (ARES-) ARES TRADING SA.
XX
PI Davids AR, Fagan RJ, Phelps CB, Power C;
XX
DR WPI; 2003-903655/82.
DR N-PSDB; ACH01277.
XX
PT New INSP052 polypeptides and nucleic acids, useful in diagnosing and
PT treating cell proliferative, autoimmune/inflammatory, cardiovascular,
PT neurological, psychiatric, developmental, genetic or metabolic disorder.
XX
PS Example 2; Fig 5; 0pp; English.
XX
CC The present invention provides the protein and coding sequences of a
CC novel human immunoglobulin domain-containing cell surface recognition
CC molecule known as INSP052. The polypeptide is useful as immunoglobulin
CC domain-containing cell surface recognition molecule. The sequences may
CC also be used in therapy or diagnosing a disease or in the manufacture of
CC a medicament for treating a disease. The disease is a cell proliferative,
CC autoimmune/inflammatory, cardiovascular, neurological, psychiatric,
CC developmental, genetic or metabolic disorder, an infection or other
CC pathological condition. The polypeptides and nucleic acids are essential
CC to the structural integrity and homeostatic functioning of most tissues.

CC The present sequence is a polypeptide shown in the invention
XX
SQ Sequence 416 AA;

Query Match 100.0%; Score 416; DB 7; Length 416;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 416; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MKRERGALSRASRALRLAPFVYLLLIQTDPLEGVNITSPVRLIHGTVGKSALLSVQYSST 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MKRERGALSRASRALRLAPFVYLLLIQTDPLEGVNITSPVRLIHGTVGKSALLSVQYSST 60

Qy         61 SSDRPVVKWQLKRDKPVTVVQSIGTEVIGTLRPDYRDRIRLFENGSLLLSDLQLADEGTY 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 SSDRPVVKWQLKRDKPVTVVQSIGTEVIGTLRPDYRDRIRLFENGSLLLSDLQLADEGTY 120

Qy        121 EVEISITDDTFTGEKTINLTVDVPISRPQVLVASTTVLELSEAFTLNCSHENGTKPSYTW 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 EVEISITDDTFTGEKTINLTVDVPISRPQVLVASTTVLELSEAFTLNCSHENGTKPSYTW 180

Qy        181 LKDGKPLLNDSRMLLSPDQKVLTITRVLMEDDDLYSCMVENPISQGRSLPVKITVYRRSS 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 LKDGKPLLNDSRMLLSPDQKVLTITRVLMEDDDLYSCMVENPISQGRSLPVKITVYRRSS 240

Qy        241 LYIILSTGGIFLLVTLVTVCACWKPSKRKQKKLEKQNSLEYMDQNDDRLKPEADTLPRSG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 LYIILSTGGIFLLVTLVTVCACWKPSKRKQKKLEKQNSLEYMDQNDDRLKPEADTLPRSG 300

Qy        301 EQERKNPMALYILKDKDSPETEENPAPEPRSATEPGPPGYSVSPAVPGRSPGLPIRSARR 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 EQERKNPMALYILKDKDSPETEENPAPEPRSATEPGPPGYSVSPAVPGRSPGLPIRSARR 360

Qy        361 YPRSPARSPATGRTHSSPPRAPSSPGRSRSASRTLRTAGVHIIREQDEAGPVEISA 416
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 YPRSPARSPATGRTHSSPPRAPSSPGRSRSASRTLRTAGVHIIREQDEAGPVEISA 416
```

RESULT 2
ABG75377
ID ABG75377 standard; protein; 416 AA.
XX
AC ABG75377;
XX
DT 22-APR-2004 (first entry)
XX
DE Human INSP052 complete protein.
XX
KW INSP052; human; cell proliferation; autoimmune disease; inflammation;
KW cardiovascular disease; neurological disease; psychiatric disease;
KW developmental disease; metabolic disorder; infection;
KW immunoglobulin domain-containing cell surface recognition molecule.
XX
OS Homo sapiens.
XX
PN WO2003093316-A2.
XX
PD 13-NOV-2003.
XX
PF 30-APR-2003; 2003WO-GB001851.
XX
PR 30-APR-2002; 2002GB-00009884.
XX
PA (ARES-) ARES TRADING SA.
XX
PI Davids AR, Fagan RJ, Phelps CB, Power C;
XX
DR WPI; 2003-903655/82.
DR N-PSDB; ACH01275.
XX
PT New INSP052 polypeptides and nucleic acids, useful in diagnosing and
PT treating cell proliferative, autoimmune/inflammatory, cardiovascular,
PT neurological, psychiatric, developmental, genetic or metabolic disorder.
XX
PS Claim 1; Page 67; 0pp; English.
XX
CC The present invention provides the protein and coding sequences of a
CC novel human immunoglobulin domain-containing cell surface recognition
CC molecule known as INSP052. The polypeptide is useful as immunoglobulin
CC domain-containing cell surface recognition molecule. The sequences may
CC also be used in therapy or diagnosing a disease or in the manufacture of
CC a medicament for treating a disease. The disease is a cell proliferative,
CC autoimmune/inflammatory, cardiovascular, neurological, psychiatric,
CC developmental, genetic or metabolic disorder, an infection or other
CC pathological condition. The polypeptides and nucleic acids are essential
CC to the structural integrity and homeostatic functioning of most tissues.
CC The present sequence is a polypeptide shown in the invention
XX
SQ Sequence 416 AA;

Query Match 100.0%; Score 416; DB 7; Length 416;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 416; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MKRERGALSRASRALRLAPFVYLLLIQTDPLEGVNITSPVRLIHGTVGKSALLSVQYSST 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MKRERGALSRASRALRLAPFVYLLLIQTDPLEGVNITSPVRLIHGTVGKSALLSVQYSST 60

Qy         61 SSDRPVVKWQLKRDKPVTVVQSIGTEVIGTLRPDYRDRIRLFENGSLLLSDLQLADEGTY 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 SSDRPVVKWQLKRDKPVTVVQSIGTEVIGTLRPDYRDRIRLFENGSLLLSDLQLADEGTY 120

Qy        121 EVEISITDDTFTGEKTINLTVDVPISRPQVLVASTTVLELSEAFTLNCSHENGTKPSYTW 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 EVEISITDDTFTGEKTINLTVDVPISRPQVLVASTTVLELSEAFTLNCSHENGTKPSYTW 180

Qy        181 LKDGKPLLNDSRMLLSPDQKVLTITRVLMEDDDLYSCMVENPISQGRSLPVKITVYRRSS 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 LKDGKPLLNDSRMLLSPDQKVLTITRVLMEDDDLYSCMVENPISQGRSLPVKITVYRRSS 240

Qy        241 LYIILSTGGIFLLVTLVTVCACWKPSKRKQKKLEKQNSLEYMDQNDDRLKPEADTLPRSG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 LYIILSTGGIFLLVTLVTVCACWKPSKRKQKKLEKQNSLEYMDQNDDRLKPEADTLPRSG 300

Qy        301 EQERKNPMALYILKDKDSPETEENPAPEPRSATEPGPPGYSVSPAVPGRSPGLPIRSARR 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 EQERKNPMALYILKDKDSPETEENPAPEPRSATEPGPPGYSVSPAVPGRSPGLPIRSARR 360

Qy        361 YPRSPARSPATGRTHSSPPRAPSSPGRSRSASRTLRTAGVHIIREQDEAGPVEISA 416
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 YPRSPARSPATGRTHSSPPRAPSSPGRSRSASRTLRTAGVHIIREQDEAGPVEISA 416
```

RESULT 3
ADO47892
ID ADO47892 standard; protein; 416 AA.
XX
AC ADO47892;
XX
DT 15-JUL-2004 (first entry)
XX
DE Human protein SEQ ID NO:9.
XX
KW human; virucide; anti-HIV; cytostatic; antiinflammatory; antiallergic;
KW immunosuppressive; antiarteriosclerotic; hypotensive; osteopathic;
KW antianaemic; neuroprotective; nootropic; antiparkinsonian; antiasthmatic;
KW haemostatic; antidiabetic; cardiant; HIV; viral infection; cancer;
KW inflammation; allergy; graft rejection; atherosclerosis; hypertension;
KW osteoporosis; anaemia; Alzheimer's disease; Parkinson's disease; asthma;
KW diabetes; myocardial infarction; haemophilia.
XX
OS Homo sapiens.
XX
PN WO2004007672-A2.
XX
PD 22-JAN-2004.
XX

PF 09-JUL-2003; 2003WO-US021703.
XX
PR 12-JUL-2002; 2002US-0395402P.
XX
PA (NUVE-) NUVELO INC.
XX
PI Rupp F, Wang J, Zhou P, Wehrman T, Wang ZW, Tang YT;
XX
DR WPI; 2004-122914/12.
DR N-PSDB; ADO47891.
XX
PT New isolated polypeptides and polynucleotides useful in diagnostics,
PT forensics, in preventing or treating diseases such as HIV and cancer, and
PT as drug targets.
XX
PS Claim 10; SEQ ID NO 9; 205pp; English.
XX
CC The invention relates to novel isolated polynucleotides and polypeptides
CC encoded by them. Also included are mutants or variants of the
CC polynucleotides and polypeptides. A polypeptide of the invention has
CC virucide, anti-HIV, cytostatic, antiinflammatory, antiallergic,
CC immunosuppressive, antiarteriosclerotic, hypotensive, osteopathic,
CC antianaemic, neuroprotective, nootropic, antiparkinsonian, antiasthmatic,
CC haemostatic, antidiabetic, and cardiant activity. The composition and
CC methods are useful in diagnostics, forensics, gene or chromosome mapping,
CC identification of mutations responsible for genetic disorders or other
CC traits, in assessing biodiversity, or in producing many other types of
CC data and products dependent on DNA and amino acid sequences. They may
CC also be used in preventing or treating diseases (e.g. HIV and other viral
CC infections, cancer, inflammation, allergies, graft rejection,
CC atherosclerosis, hypertension, osteoporosis, anaemia, Alzheimer's
CC disease, Parkinson's disease, asthma, diabetes, myocardial infarction or
CC haemophilia). They may also be used as targets in drug screening. The
CC present sequence represents a polypeptide of the invention.
XX
SQ Sequence 416 AA;

Query Match 100.0%; Score 416; DB 8; Length 416;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 416; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MKRERGALSRASRALRLAPFVYLLLIQTDPLEGVNITSPVRLIHGTVGKSALLSVQYSST 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MKRERGALSRASRALRLAPFVYLLLIQTDPLEGVNITSPVRLIHGTVGKSALLSVQYSST 60

Qy         61 SSDRPVVKWQLKRDKPVTVVQSIGTEVIGTLRPDYRDRIRLFENGSLLLSDLQLADEGTY 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 SSDRPVVKWQLKRDKPVTVVQSIGTEVIGTLRPDYRDRIRLFENGSLLLSDLQLADEGTY 120

Qy        121 EVEISITDDTFTGEKTINLTVDVPISRPQVLVASTTVLELSEAFTLNCSHENGTKPSYTW 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 EVEISITDDTFTGEKTINLTVDVPISRPQVLVASTTVLELSEAFTLNCSHENGTKPSYTW 180

Qy        181 LKDGKPLLNDSRMLLSPDQKVLTITRVLMEDDDLYSCMVENPISQGRSLPVKITVYRRSS 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 LKDGKPLLNDSRMLLSPDQKVLTITRVLMEDDDLYSCMVENPISQGRSLPVKITVYRRSS 240

Qy        241 LYIILSTGGIFLLVTLVTVCACWKPSKRKQKKLEKQNSLEYMDQNDDRLKPEADTLPRSG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 LYIILSTGGIFLLVTLVTVCACWKPSKRKQKKLEKQNSLEYMDQNDDRLKPEADTLPRSG 300

Qy        301 EQERKNPMALYILKDKDSPETEENPAPEPRSATEPGPPGYSVSPAVPGRSPGLPIRSARR 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 EQERKNPMALYILKDKDSPETEENPAPEPRSATEPGPPGYSVSPAVPGRSPGLPIRSARR 360

Qy        361 YPRSPARSPATGRTHSSPPRAPSSPGRSRSASRTLRTAGVHIIREQDEAGPVEISA 416
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 YPRSPARSPATGRTHSSPPRAPSSPGRSRSASRTLRTAGVHIIREQDEAGPVEISA 416
```

RESULT 4
ADS11056
ID ADS11056 standard; protein; 416 AA.
XX
AC ADS11056;
XX
DT 16-DEC-2004 (first entry)
XX
DE Human therapeutic protein - SEQ ID 1293.
XX
KW antiinflammatory; neuroprotective; antianaemic; cytostatic; vulnerary;
KW inflammatory; haematopoiesis; immunity; neurodegenerative; stem cell;
KW aplastic anaemia; cancer; wound healing; gene therapy.
XX
OS Homo sapiens.
XX
PN WO2004080148-A2.
XX
PD 23-SEP-2004.
XX
PF 30-SEP-2003; 2003WO-US030720.
XX
PR 02-OCT-2002; 2002US-0416186P.
XX
PA (NUVE-) NUVELO INC.
XX
PI Tang YT, Asundi V, Ren F, Zhang J, Wehrman T, Wang Z, Ma Y;
PI Wang D, Chen R, Zhao QA, Wang J, Ghosh M, Xue AJ, Weng G, Zhou P;
XX
DR WPI; 2004-668857/65.
DR N-PSDB; ADS10372.
XX
PT New polynucleotide, useful in preparing a composition for diagnosing or
PT treating inflammatory, neurodegenerative or stem cell disorders, e.g.,
PT aplastic anemia or cancer for promoting wound healing.
XX
PS Claim 20; SEQ ID NO 1293; 718pp; English.
XX
CC The invention relates to a novel isolated polynucleotide and the encoded
CC polypeptide. The molecules of the invention demonstrate antiinflammatory,
CC neuroprotective, antianaemic, cytostatic and vulnerary activities and may
CC be useful in preparing a composition for diagnosing or treating
CC inflammatory, haematopoietic, immune, neurodegenerative or stem cell
CC disorders, such as aplastic anaemia or cancer, as well as for promoting
CC wound healing. The molecules may also be utilised during gene therapy
CC procedures. The current sequence is that of a human therapeutic protein
CC of the invention. The current sequence is not shown explicitly within the
CC specification but can be accessed from the WIPO web-site.
XX
SQ Sequence 416 AA;

Query Match 100.0%; Score 416; DB 8; Length 416;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 416; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MKRERGALSRASRALRLAPFVYLLLIQTDPLEGVNITSPVRLIHGTVGKSALLSVQYSST 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MKRERGALSRASRALRLAPFVYLLLIQTDPLEGVNITSPVRLIHGTVGKSALLSVQYSST 60

Qy         61 SSDRPVVKWQLKRDKPVTVVQSIGTEVIGTLRPDYRDRIRLFENGSLLLSDLQLADEGTY 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 SSDRPVVKWQLKRDKPVTVVQSIGTEVIGTLRPDYRDRIRLFENGSLLLSDLQLADEGTY 120

Qy        121 EVEISITDDTFTGEKTINLTVDVPISRPQVLVASTTVLELSEAFTLNCSHENGTKPSYTW 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 EVEISITDDTFTGEKTINLTVDVPISRPQVLVASTTVLELSEAFTLNCSHENGTKPSYTW 180

Qy        181 LKDGKPLLNDSRMLLSPDQKVLTITRVLMEDDDLYSCMVENPISQGRSLPVKITVYRRSS 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 LKDGKPLLNDSRMLLSPDQKVLTITRVLMEDDDLYSCMVENPISQGRSLPVKITVYRRSS 240

Qy        241 LYIILSTGGIFLLVTLVTVCACWKPSKRKQKKLEKQNSLEYMDQNDDRLKPEADTLPRSG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 LYIILSTGGIFLLVTLVTVCACWKPSKRKQKKLEKQNSLEYMDQNDDRLKPEADTLPRSG 300

Qy        301 EQERKNPMALYILKDKDSPETEENPAPEPRSATEPGPPGYSVSPAVPGRSPGLPIRSARR 360
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 EQERKNPMALYILKDKDSPETEENPAPEPRSATEPGPPGYSVSPAVPGRSPGLPIRSARR 360

Qy        361 YPRSPARSPATGRTHSSPPRAPSSPGRSRSASRTLRTAGVHIIREQDEAGPVEISA 416
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 YPRSPARSPATGRTHSSPPRAPSSPGRSRSASRTLRTAGVHIIREQDEAGPVEISA 416


RESULT 5
ADO47895
ID   ADO47895 standard; protein; 383 AA.
XX
AC   ADO47895;
XX
DT   15-JUL-2004  (first entry)
XX
DE   Human mature protein SEQ ID NO:12.
XX
KW   human; virucide; anti-HIV; cytostatic; antiinflammatory; antiallergic;
KW   immunosuppressive; antiarteriosclerotic; hypotensive; osteopathic;
KW   antianaemic; neuroprotective; nootropic; antiparkinsonian; antiasthmatic;
KW   haemostatic; antidiabetic; cardiant; HIV; viral infection; cancer;
KW   inflammation; allergy; graft rejection; atherosclerosis; hypertension;
KW   osteoporosis; anaemia; Alzheimer's disease; Parkinson's disease; asthma;
KW   diabetes; myocardial infarction; haemophilia.
XX
OS   Homo sapiens.
XX
PN   WO2004007672-A2.
XX
PD   22-JAN-2004.
XX
PF   09-JUL-2003; 2003WO-US021703.
XX
PR   12-JUL-2002; 2002US-0395402P.
XX
PA   (NUVE-) NUVELO INC.
XX
PI   Rupp F,  Wang J,  Zhou P,  Wehrman T,  Wang ZW,  Tang YT;
XX
DR   WPI; 2004-122914/12.
DR   N-PSDB; ADO47893.
XX
PT   New isolated polypeptides and polynucleotides useful in diagnostics,
PT   forensics, in preventing or treating diseases such as HIV and cancer, and
PT   as drug targets.
XX
PS   Claim 10; SEQ ID NO 12; 205pp; English.
XX
CC   The invention relates to novel isolated polynucleotides and polypeptides
CC   encoded by them. Also included are mutants or variants of the
CC   polynucleotides and polypeptides. A polypeptide of the invention has
CC   virucide, anti-HIV, cytostatic, antiinflammatory, antiallergic,
CC   immunosuppressive, antiarteriosclerotic, hypotensive, osteopathic,
CC   antianaemic, neuroprotective, nootropic, antiparkinsonian, antiasthmatic,
CC   haemostatic, antidiabetic, and cardiant activity. The composition and
CC   methods are useful in diagnostics, forensics, gene or chromosome mapping,
CC   identification of mutations responsible for genetic disorders or other
CC   traits, in assessing biodiversity, or in producing many other types of
CC   data and products dependent on DNA and amino acid sequences. They may
CC   also be used in preventing or treating diseases (e.g. HIV and other viral
CC   infections, cancer, inflammation, allergies, graft rejection,
CC   atherosclerosis, hypertension, osteoporosis, anaemia, Alzheimer's
CC   disease, Parkinson's disease, asthma, diabetes, myocardial infarction or
CC   haemophilia). They may also be used as targets in drug screening. The
CC   present sequence represents a polypeptide of the invention.
XX
SQ   Sequence 383 AA;

  Query Match             92.1%;  Score 383;  DB 8;  Length 383;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  383;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         34 VNITSPVRLIHGTVGKSALLSVQYSSTSSDRPVVKWQLKRDKPVTVVQSIGTEVIGTLRP 93
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 VNITSPVRLIHGTVGKSALLSVQYSSTSSDRPVVKWQLKRDKPVTVVQSIGTEVIGTLRP 60

Qy         94 DYRDRIRLFENGSLLLSDLQLADEGTYEVEISITDDTFTGEKTINLTVDVPISRPQVLVA 153
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 DYRDRIRLFENGSLLLSDLQLADEGTYEVEISITDDTFTGEKTINLTVDVPISRPQVLVA 120

Qy        154 STTVLELSEAFTLNCSHENGTKPSYTWLKDGKPLLNDSRMLLSPDQKVLTITRVLMEDDD 213
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 STTVLELSEAFTLNCSHENGTKPSYTWLKDGKPLLNDSRMLLSPDQKVLTITRVLMEDDD 180

Qy        214 LYSCMVENPISQGRSLPVKITVYRRSSLYIILSTGGIFLLVTLVTVCACWKPSKRKQKKL 273
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 LYSCMVENPISQGRSLPVKITVYRRSSLYIILSTGGIFLLVTLVTVCACWKPSKRKQKKL 240

Qy        274 EKQNSLEYMDQNDDRLKPEADTLPRSGEQERKNPMALYILKDKDSPETEENPAPEPRSAT 333
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 EKQNSLEYMDQNDDRLKPEADTLPRSGEQERKNPMALYILKDKDSPETEENPAPEPRSAT 300

Qy        334 EPGPPGYSVSPAVPGRSPGLPIRSARRYPRSPARSPATGRTHSSPPRAPSSPGRSRSASR 393
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 EPGPPGYSVSPAVPGRSPGLPIRSARRYPRSPARSPATGRTHSSPPRAPSSPGRSRSASR 360

Qy        394 TLRTAGVHIIREQDEAGPVEISA 416
                |||||||||||||||||||||||
Db        361 TLRTAGVHIIREQDEAGPVEISA 383


RESULT 6
AAE14784
ID   AAE14784 standard; protein; 298 AA.
XX
AC   AAE14784;
XX
DT   30-OCT-2002  (first entry)
XX
DE   Human immunoglobulin superfamily protein (IGSFP)-4.
XX
KW   Human; immunoglobulin superfamily protein-4; IGSFP-4; asthma;
KW   immune system disorder; acquired immune deficiency syndrome; AIDS;
KW   atherosclerosis; neurological disorder; Alzheimer's disease;
KW   Parkinson's disease; developmental disorder; renal tubular acidosis;
KW   anaemia; muscle disorder; cardiomyopathy; myocarditis; cancer;
KW   cell proliferative disorder; arteriosclerosis; hepatitis.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1. .33
FT                   /label= Signal_peptide
FT   Protein         34. .298
FT                   /note= "Mature IGSFP-4"
FT   Region          43. .231
FT                   /note= "Antigen precursor signal immunoglobulin fold
FT                   glycoprotein T cell surface transmembrane"
FT   Domain          48. .120
FT                   /label= Immunoglobulin_domain
FT   Domain          161. .219
FT                   /label= Immunoglobulin_domain
FT   Domain          243. .263
FT                   /label= Transmembrane_domain
XX
PN   WO200240671-A2.
XX
PD   23-MAY-2002.
XX
PF   15-NOV-2001; 2001WO-US044974.
XX
PR   16-NOV-2000; 2000US-0249645P.
XX
PA   (INCY-) INCYTE GENOMICS INC.
XX
```

PI Baughn MR, Lu DAM, Yue H, Elliott VS, Thangavelu K, Ramkumar J;
PI Lu Y, Lo TP, Gururajan R, Gandhi AR, Arvizu C, Yao MG;
XX
DR WPI; 2002-519384/55.
DR N-PSDB; AAD36780.
XX
PT Novel human immunoglobulin superfamily polypeptide, useful in diagnosis,
PT prevention or treatment of immune system, neurological, developmental,
PT muscle and cell proliferative disorders.
XX
PS Claim 1; Page 109-110; 122pp; English.
XX
CC The present sequence is human immunoglobulin superfamily protein (IGSFP)-
CC 4. The IGSFP polypeptide and polynucleotide are useful for diagnosing,
CC treating or preventing disorders associated with aberrant expression of
CC IGSFP e.g. immune system disorders (e.g. acquired immune deficiency
CC syndrome (AIDS), asthma, atherosclerosis, psoriasis, uveitis),
CC neurological disorders (e.g. Alzheimer's disease, Huntington's disease,
CC dementia, Parkinson's disease), developmental disorders (e.g. renal
CC tubular acidosis, epilepsy, anaemia), muscle disorders (e.g.
CC cardiomyopathy, myocarditis), or cell proliferative disorders (e.g.
CC arteriosclerosis, cirrhosis, hepatitis, and cancer). The polypeptide and
CC polynucleotide are also useful for assessing the effects of exogenous
CC compounds on their expression. The polypeptide is useful in drug
CC screening techniques, to analyse the proteome of a tissue or cell type,
CC as elements on a microarray. The polynucleotide is useful for creating
CC knock-in humanised animals or transgenic animals to model human diseases,
CC in somatic or germline gene therapy, to generate a transcript image of a
CC tissue or cell type, for detecting differences in the chromosomal
CC location due to translocation, inversion among normal, carrier or
CC affected individuals, and as hybridisation probes for mapping naturally
CC occurring genomic sequences
XX
SQ Sequence 298 AA;

Query Match 70.0%; Score 291; DB 5; Length 298;
Best Local Similarity 100.0%; Pred. No. 4.5e-274;
Matches 291; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MKRERGALSRASRALRLAPFVYLLLIQTDPLEGVNITSPVRLIHGTVGKSALLSVQYSST 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MKRERGALSRASRALRLAPFVYLLLIQTDPLEGVNITSPVRLIHGTVGKSALLSVQYSST 60

Qy         61 SSDRPVVKWQLKRDKPVTVVQSIGTEVIGTLRPDYRDRIRLFENGSLLLSDLQLADEGTY 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 SSDRPVVKWQLKRDKPVTVVQSIGTEVIGTLRPDYRDRIRLFENGSLLLSDLQLADEGTY 120

Qy        121 EVEISITDDTFTGEKTINLTVDVPISRPQVLVASTTVLELSEAFTLNCSHENGTKPSYTW 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 EVEISITDDTFTGEKTINLTVDVPISRPQVLVASTTVLELSEAFTLNCSHENGTKPSYTW 180

Qy        181 LKDGKPLLNDSRMLLSPDQKVLTITRVLMEDDDLYSCMVENPISQGRSLPVKITVYRRSS 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 LKDGKPLLNDSRMLLSPDQKVLTITRVLMEDDDLYSCMVENPISQGRSLPVKITVYRRSS 240

Qy        241 LYIILSTGGIFLLVTLVTVCACWKPSKRKQKKLEKQNSLEYMDQNDDRLKP 291
              |||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 LYIILSTGGIFLLVTLVTVCACWKPSKRKQKKLEKQNSLEYMDQNDDRLKP 291
```

RESULT 7
ADO47887
ID ADO47887 standard; protein; 270 AA.
XX
AC ADO47887;
XX
DT 15-JUL-2004 (first entry)
XX
DE Human protein SEQ ID NO:4.
XX
KW human; virucide; anti-HIV; cytostatic; antiinflammatory; antiallergic;
KW immunosuppressive; antiarteriosclerotic; hypotensive; osteopathic;
KW antianaemic; neuroprotective; nootropic; antiparkinsonian; antiasthmatic;
KW haemostatic; antidiabetic; cardiant; HIV; viral infection; cancer;
KW inflammation; allergy; graft rejection; atherosclerosis; hypertension;
KW osteoporosis; anaemia; Alzheimer's disease; Parkinson's disease; asthma;
KW diabetes; myocardial infarction; haemophilia.
XX
OS Homo sapiens.
XX
PN WO2004007672-A2.
XX
PD 22-JAN-2004.
XX
PF 09-JUL-2003; 2003WO-US021703.
XX
PR 12-JUL-2002; 2002US-0395402P.
XX
PA (NUVE-) NUVELO INC.
XX
PI Rupp F, Wang J, Zhou P, Wehrman T, Wang ZW, Tang YT;
XX
DR WPI; 2004-122914/12.
DR N-PSDB; ADO47886.
XX
PT New isolated polypeptides and polynucleotides useful in diagnostics,
PT forensics, in preventing or treating diseases such as HIV and cancer, and
PT as drug targets.
XX
PS Claim 10; SEQ ID NO 4; 205pp; English.
XX
CC The invention relates to novel isolated polynucleotides and polypeptides
CC encoded by them. Also included are mutants or variants of the
CC polynucleotides and polypeptides. A polypeptide of the invention has
CC virucide, anti-HIV, cytostatic, antiinflammatory, antiallergic,
CC immunosuppressive, antiarteriosclerotic, hypotensive, osteopathic,
CC antianaemic, neuroprotective, nootropic, antiparkinsonian, antiasthmatic,
CC haemostatic, antidiabetic, and cardiant activity. The composition and
CC methods are useful in diagnostics, forensics, gene or chromosome mapping,
CC identification of mutations responsible for genetic disorders or other
CC traits, in assessing biodiversity, or in producing many other types of
CC data and products dependent on DNA and amino acid sequences. They may
CC also be used in preventing or treating diseases (e.g. HIV and other viral
CC infections, cancer, inflammation, allergies, graft rejection,
CC atherosclerosis, hypertension, osteoporosis, anaemia, Alzheimer's
CC disease, Parkinson's disease, asthma, diabetes, myocardial infarction or
CC haemophilia). They may also be used as targets in drug screening. The
CC present sequence represents a polypeptide of the invention.
XX
SQ Sequence 270 AA;

Query Match 64.4%; Score 268; DB 8; Length 270;
Best Local Similarity 100.0%; Pred. No. 9.9e-252;
Matches 268; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MKRERGALSRASRALRLAPFVYLLLIQTDPLEGVNITSPVRLIHGTVGKSALLSVQYSST 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MKRERGALSRASRALRLAPFVYLLLIQTDPLEGVNITSPVRLIHGTVGKSALLSVQYSST 60

Qy         61 SSDRPVVKWQLKRDKPVTVVQSIGTEVIGTLRPDYRDRIRLFENGSLLLSDLQLADEGTY 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 SSDRPVVKWQLKRDKPVTVVQSIGTEVIGTLRPDYRDRIRLFENGSLLLSDLQLADEGTY 120

Qy        121 EVEISITDDTFTGEKTINLTVDVPISRPQVLVASTTVLELSEAFTLNCSHENGTKPSYTW 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 EVEISITDDTFTGEKTINLTVDVPISRPQVLVASTTVLELSEAFTLNCSHENGTKPSYTW 180

Qy        181 LKDGKPLLNDSRMLLSPDQKVLTITRVLMEDDDLYSCMVENPISQGRSLPVKITVYRRSS 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 LKDGKPLLNDSRMLLSPDQKVLTITRVLMEDDDLYSCMVENPISQGRSLPVKITVYRRSS 240

Qy        241 LYIILSTGGIFLLVTLVTVCACWKPSKR 268
              ||||||||||||||||||||||||||||
Db        241 LYIILSTGGIFLLVTLVTVCACWKPSKR 268
```

```
RESULT 8
ADS11055
ID   ADS11055 standard; protein; 270 AA.
XX
AC   ADS11055;
XX
DT   16-DEC-2004  (first entry)
XX
DE   Human therapeutic protein - SEQ ID 1292.
XX
KW   antiinflammatory; neuroprotective; antianaemic; cytostatic; vulnerary;
KW   inflammatory; haematopoiesis; immunity; neurodegenerative; stem cell;
KW   aplastic anaemia; cancer; wound healing; gene therapy.
XX
OS   Homo sapiens.
XX
PN   WO2004080148-A2.
XX
PD   23-SEP-2004.
XX
PF   30-SEP-2003; 2003WO-US030720.
XX
PR   02-OCT-2002; 2002US-0416186P.
XX
PA   (NUVE-) NUVELO INC.
XX
PI   Tang YT,  Asundi V,  Ren F,  Zhang J,  Wehrman T,  Wang Z,  Ma Y;
PI   Wang D,  Chen R,  Zhao QA,  Wang J,  Ghosh M,  Xue AJ,  Weng G,  Zhou P;
XX
DR   WPI; 2004-668857/65.
DR   N-PSDB; ADS10371.
XX
PT   New polynucleotide, useful in preparing a composition for diagnosing or
PT   treating inflammatory, neurodegenerative or stem cell disorders, e.g.,
PT   aplastic anemia or cancer for promoting wound healing.
XX
PS   Claim 20; SEQ ID NO 1292; 718pp; English.
XX
CC   The invention relates to a novel isolated polynucleotide and the encoded
CC   polypeptide. The molecules of the invention demonstrate antiinflammatory,
CC   neuroprotective, antianaemic, cytostatic and vulnerary activities and may
CC   be useful in preparing a composition for diagnosing or treating
CC   inflammatory, haematopoietic, immune, neurodegenerative or stem cell
CC   disorders, such as aplastic anaemia or cancer, as well as for promoting
CC   wound healing. The molecules may also be utilised during gene therapy
CC   procedures. The current sequence is that of a human therapeutic protein
CC   of the invention. The current sequence is not shown explicitly within the
CC   specification but can be accessed from the WIPO web-site.
XX
SQ   Sequence 270 AA;

  Query Match             64.4%;  Score 268;  DB 8;  Length 270;
  Best Local Similarity   100.0%;  Pred. No. 9.9e-252;
  Matches  268;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MKRERGALSRASRALRLAPFVYLLLIQTDPLEGVNITSPVRLIHGTVGKSALLSVQYSST 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MKRERGALSRASRALRLAPFVYLLLIQTDPLEGVNITSPVRLIHGTVGKSALLSVQYSST 60

Qy         61 SSDRPVVKWQLKRDKPVTVVQSIGTEVIGTLRPDYRDRIRLFENGSLLLSDLQLADEGTY 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 SSDRPVVKWQLKRDKPVTVVQSIGTEVIGTLRPDYRDRIRLFENGSLLLSDLQLADEGTY 120

Qy        121 EVEISITDDTFTGEKTINLTVDVPISRPQVLVASTTVLELSEAFTLNCSHENGTKPSYTW 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 EVEISITDDTFTGEKTINLTVDVPISRPQVLVASTTVLELSEAFTLNCSHENGTKPSYTW 180

Qy        181 LKDGKPLLNDSRMLLSPDQKVLTITRVLMEDDDLYSCMVENPISQGRSLPVKITVYRRSS 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 LKDGKPLLNDSRMLLSPDQKVLTITRVLMEDDDLYSCMVENPISQGRSLPVKITVYRRSS 240

Qy        241 LYIILSTGGIFLLVTLVTVCACWKPSKR 268
              ||||||||||||||||||||||||||||
Db        241 LYIILSTGGIFLLVTLVTVCACWKPSKR 268


RESULT 9
ABG75380
ID   ABG75380 standard; protein; 246 AA.
XX
AC   ABG75380;
XX
DT   22-APR-2004  (first entry)
XX
DE   INSP052 extracellular domain protein.
XX
KW   INSP052; human; cell proliferation; autoimmune disease; inflammation;
KW   cardiovascular disease; neurological disease; psychiatric disease;
KW   developmental disease; metabolic disorder; infection;
KW   immunoglobulin domain-containing cell surface recognition molecule.
XX
OS   Unidentified.
XX
PN   WO2003093316-A2.
XX
PD   13-NOV-2003.
XX
PF   30-APR-2003; 2003WO-GB001851.
XX
PR   30-APR-2002; 2002GB-00009884.
XX
PA   (ARES-) ARES TRADING SA.
XX
PI   Davids AR,  Fagan RJ,  Phelps CB,  Power C;
XX
DR   WPI; 2003-903655/82.
DR   N-PSDB; ACH01279.
XX
PT   New INSP052 polypeptides and nucleic acids, useful in diagnosing and
PT   treating cell proliferative, autoimmune/inflammatory, cardiovascular,
PT   neurological, psychiatric, developmental, genetic or metabolic disorder.
XX
PS   Claim 1; Fig 7; 0pp; English.
XX
CC   The present invention provides the protein and coding sequences of a
CC   novel human immunoglobulin domain-containing cell surface recognition
CC   molecule known as INSP052. The polypeptide is useful as immunoglobulin
CC   domain-containing cell surface recognition molecule. The sequences may
CC   also be used in therapy or diagnosing a disease or in the manufacture of
CC   a medicament for treating a disease. The disease is a cell proliferative,
CC   autoimmune/inflammatory, cardiovascular, neurological, psychiatric,
CC   developmental, genetic or metabolic disorder, an infection or other
CC   pathological condition. The polypeptides and nucleic acids are essential
CC   to the structural integrity and homeostatic functioning of most tissues.
CC   The present sequence is a polypeptide shown in the invention
XX
SQ   Sequence 246 AA;

  Query Match             57.7%;  Score 240;  DB 7;  Length 246;
  Best Local Similarity   100.0%;  Pred. No. 1.6e-224;
  Matches  240;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MKRERGALSRASRALRLAPFVYLLLIQTDPLEGVNITSPVRLIHGTVGKSALLSVQYSST 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MKRERGALSRASRALRLAPFVYLLLIQTDPLEGVNITSPVRLIHGTVGKSALLSVQYSST 60

Qy         61 SSDRPVVKWQLKRDKPVTVVQSIGTEVIGTLRPDYRDRIRLFENGSLLLSDLQLADEGTY 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 SSDRPVVKWQLKRDKPVTVVQSIGTEVIGTLRPDYRDRIRLFENGSLLLSDLQLADEGTY 120

Qy        121 EVEISITDDTFTGEKTINLTVDVPISRPQVLVASTTVLELSEAFTLNCSHENGTKPSYTW 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 EVEISITDDTFTGEKTINLTVDVPISRPQVLVASTTVLELSEAFTLNCSHENGTKPSYTW 180

Qy        181 LKDGKPLLNDSRMLLSPDQKVLTITRVLMEDDDLYSCMVENPISQGRSLPVKITVYRRSS 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        181 LKDGKPLLNDSRMLLSPDQKVLTITRVLMEDDDLYSCMVENPISQGRSLPVKITVYRRSS 240

RESULT 10
ADO47890
ID   ADO47890 standard; protein; 237 AA.
XX
AC   ADO47890;
XX
DT   15-JUL-2004  (first entry)
XX
DE   Human mature protein SEQ ID NO:7.
XX
KW   human; virucide; anti-HIV; cytostatic; antiinflammatory; antiallergic;
KW   immunosuppressive; antiarteriosclerotic; hypotensive; osteopathic;
KW   antianaemic; neuroprotective; nootropic; antiparkinsonian; antiasthmatic;
KW   haemostatic; antidiabetic; cardiant; HIV; viral infection; cancer;
KW   inflammation; allergy; graft rejection; atherosclerosis; hypertension;
KW   osteoporosis; anaemia; Alzheimer's disease; Parkinson's disease; asthma;
KW   diabetes; myocardial infarction; haemophilia.
XX
OS   Homo sapiens.
XX
PN   WO2004007672-A2.
XX
PD   22-JAN-2004.
XX
PF   09-JUL-2003; 2003WO-US021703.
XX
PR   12-JUL-2002; 2002US-0395402P.
XX
PA   (NUVE-) NUVELO INC.
XX
PI   Rupp F,  Wang J,  Zhou P,  Wehrman T,  Wang ZW,  Tang YT;
XX
DR   WPI; 2004-122914/12.
DR   N-PSDB; ADO47888.
XX
PT   New isolated polypeptides and polynucleotides useful in diagnostics,
PT   forensics, in preventing or treating diseases such as HIV and cancer, and
PT   as drug targets.
XX
PS   Claim 10; SEQ ID NO 7; 205pp; English.
XX
CC   The invention relates to novel isolated polynucleotides and polypeptides
CC   encoded by them. Also included are mutants or variants of the
CC   polynucleotides and polypeptides. A polypeptide of the invention has
CC   virucide, anti-HIV, cytostatic, antiinflammatory, antiallergic,
CC   immunosuppressive, antiarteriosclerotic, hypotensive, osteopathic,
CC   antianaemic, neuroprotective, nootropic, antiparkinsonian, antiasthmatic,
CC   haemostatic, antidiabetic, and cardiant activity. The composition and
CC   methods are useful in diagnostics, forensics, gene or chromosome mapping,
CC   identification of mutations responsible for genetic disorders or other
CC   traits, in assessing biodiversity, or in producing many other types of
CC   data and products dependent on DNA and amino acid sequences. They may
CC   also be used in preventing or treating diseases (e.g. HIV and other viral
CC   infections, cancer, inflammation, allergies, graft rejection,
CC   atherosclerosis, hypertension, osteoporosis, anaemia, Alzheimer's
CC   disease, Parkinson's disease, asthma, diabetes, myocardial infarction or
CC   haemophilia). They may also be used as targets in drug screening. The
CC   present sequence represents a polypeptide of the invention.
XX
SQ   Sequence 237 AA;

  Query Match             56.5%;  Score 235;  DB 8;  Length 237;
  Best Local Similarity   100.0%;  Pred. No. 1.1e-219;
  Matches  235;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         34 VNITSPVRLIHGTVGKSALLSVQYSSTSSDRPVVKWQLKRDKPVTVVQSIGTEVIGTLRP 93
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 VNITSPVRLIHGTVGKSALLSVQYSSTSSDRPVVKWQLKRDKPVTVVQSIGTEVIGTLRP 60

Qy         94 DYRDRIRLFENGSLLLSDLQLADEGTYEVEISITDDTFTGEKTINLTVDVPISRPQVLVA 153
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 DYRDRIRLFENGSLLLSDLQLADEGTYEVEISITDDTFTGEKTINLTVDVPISRPQVLVA 120

Qy        154 STTVLELSEAFTLNCSHENGTKPSYTWLKDGKPLLNDSRMLLSPDQKVLTITRVLMEDDD 213
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 STTVLELSEAFTLNCSHENGTKPSYTWLKDGKPLLNDSRMLLSPDQKVLTITRVLMEDDD 180

Qy        214 LYSCMVENPISQGRSLPVKITVYRRSSLYIILSTGGIFLLVTLVTVCACWKPSKR 268
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 LYSCMVENPISQGRSLPVKITVYRRSSLYIILSTGGIFLLVTLVTVCACWKPSKR 235

RESULT 11
AAE26421
ID   AAE26421 standard; protein; 224 AA.
XX
AC   AAE26421;
XX
DT   13-DEC-2002  (first entry)
XX
DE   Human transmembrane protein (TMP)-7 protein.
XX
KW   Human; transmembrane protein; TMP-7; developmental disorder; epilepsy;
KW   prostatitis; infertility; neurological disorder; Alzheimer's disease;
KW   anaemia; stroke; cardiovascular disorder; hypertension; atherosclerosis;
KW   gastrointestinal disorder; anorexia; Crohn's disease; lipid metabolism;
KW   hypercholesterolaemia; hyperlipidaemia; cell proliferative disorder;
KW   psoriasis; autoimmune disorder; acquired immune deficiency syndrome;
KW   AIDS; cancer; gout; Grave's disease; transgenic; transgenic animal;
KW   gene therapy; antiinfertility; anticonvulsant; hypotensive; nootropic;
KW   neuroprotective; cerebroprotective; antiinflammatory; cytostatic;
KW   antithyroid.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Domain          51. .71
FT                   /note= "Transmembrane domain"
XX
PN   WO200234783-A2.
XX
PD   02-MAY-2002.
XX
PF   26-OCT-2001; 2001WO-US049670.
XX
PR   27-OCT-2000; 2000US-0244017P.
PR   22-NOV-2000; 2000US-0252855P.
PR   07-DEC-2000; 2000US-0251825P.
PR   12-DEC-2000; 2000US-0255085P.
XX
PA   (INCY-) INCYTE GENOMICS INC.
XX
PI   Warren BA,  Xu Y,  Yue H,  Batra S,  Burford N,  Gandhi AR,  Walia NK;
PI   Arvizu C,  Tang YT,  Lu DAM,  Duggan BM,  Baughn MR,  Lee EA,  Khan FA;
PI   Nguyen DB,  Azimzai Y,  Yao MG,  Lal PG,  Thangavelu K,  Ramkumar J;
PI   Tran B,  Ding L,  Au-Young J;
XX
DR   WPI; 2002-463354/49.
DR   N-PSDB; AAD44098.
XX
PT   Novel human transmembrane proteins and polynucleotides useful for
PT   diagnosing, treating or preventing infertility, anemia, hypertension,
PT   anorexia, hypercholesterolemia, cancer, gout, Grave's disease.
XX
PS   Claim 62; Page 132-133; 163pp; English.
XX
CC   The present invention relates to novel human transmembrane proteins (TMP)
CC   and polynucleotides encoding such proteins. Sequences of the invention
CC   are useful for treating diseases or conditions associated with abnormal
CC   expression of TMP such as disorders of reproduction (e.g. infertility,
CC   prostatitis), developmental (e.g. anaemia, epilepsy), gastrointestinal
CC   (e.g. anorexia, Crohn's disease), neurological (e.g. Alzheimer's disease,
CC   stroke), lipid metabolism (e.g. hypercholesterolaemia, hyperlipidaemia),
```

```
CC   cardiovascular (e.g. atherosclerosis, hypertension), cell proliferative
CC   (e.g. cancer, psoriasis) and autoimmune disorders (e.g. acquired immune
CC   deficiency syndrome (AIDS), gout, Grave's disease). They are useful for
CC   creating knockout humanised animals or transgenic animals to model human
CC   disease. Sequences of the invention are also used in gene therapy. The
CC   present sequence is TMP-7 protein
XX
SQ   Sequence 224 AA;

  Query Match             53.8%;  Score 224;  DB 5;  Length 224;
  Best Local Similarity   100.0%;  Pred. No. 5.3e-209;
  Matches  224;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        193 MLLSPDQKVLTITRVLMEDDDLYSCMVENPISQGRSLPVKITVYRRSSLYIILSTGGIFL 252
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MLLSPDQKVLTITRVLMEDDDLYSCMVENPISQGRSLPVKITVYRRSSLYIILSTGGIFL 60

Qy        253 LVTLVTVCACWKPSKRKQKKLEKQNSLEYMDQNDDRLKPEADTLPRSGEQERKNPMALYI 312
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LVTLVTVCACWKPSKRKQKKLEKQNSLEYMDQNDDRLKPEADTLPRSGEQERKNPMALYI 120

Qy        313 LKDKDSPETEENPAPEPRSATEPGPPGYSVSPAVPGRSPGLPIRSARRYPRSPARSPATG 372
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 LKDKDSPETEENPAPEPRSATEPGPPGYSVSPAVPGRSPGLPIRSARRYPRSPARSPATG 180

Qy        373 RTHSSPPRAPSSPGRSRSASRTLRTAGVHIIREQDEAGPVEISA 416
              |||||||||||||||||||||||||||||||||||||||||||||
Db        181 RTHSSPPRAPSSPGRSRSASRTLRTAGVHIIREQDEAGPVEISA 224


RESULT 12
ADM87341
ID   ADM87341 standard; protein; 256 AA.
XX
AC   ADM87341;
XX
DT   03-JUN-2004  (first entry)
XX
DE   Human protein SEQ ID NO:434.
XX
KW   respiratory; cytostatic; antiarthritic; antiinflammatory;
KW   gastrointestinal; antibacterial; immunosuppressive; antidiabetic;
KW   antirheumatic; gene therapy; molecular weight marker; chromosome marker;
KW   chromosome tag; genetic fingerprinting; nutritional supplement; cancer;
KW   inflammatory condition; arthritis; inflammatory bowel disease;
KW   Crohn's disease; sepsis; rheumatoid arthritis; diabetes mellitus type 1;
KW   graft versus host disease; human.
XX
OS   Homo sapiens.
XX
PN   WO2004009834-A2.
XX
PD   29-JAN-2004.
XX
PF   19-JUL-2002; 2002WO-US022858.
XX
PR   21-JUL-2001; 2001US-0306971P.
PR   28-MAR-2002; 2002US-00112944.
XX
PA   (NUVE-) NUVELO INC.
XX
PI   Tang YT,  Yang Y,  Weng G,  Zhang J,  Ren F,  Xue A,  Wang J;
PI   Wehrman T,  Ghosh MJ,  Wang D,  Zhao QA,  Wang Z;
XX
DR   WPI; 2004-143291/14.
DR   N-PSDB; ADM87097.
XX
PT   New isolated polynucleotides and polypeptides, useful for treating, e.g.
PT   cancer, lung or liver fibrosis, arthritis, inflammatory bowel disease,
PT   Crohn's disease, rheumatoid arthritis, diabetes mellitus type 1 or graft
PT   versus host disease.
XX
PS   Claim 20; SEQ ID NO 434; 591pp; English.
XX
CC   The present invention describes an isolated polynucleotide (I): (a)
CC   comprising a nucleotide sequence selected from SEQ ID NO:1-244; or (b)
CC   which encodes a polypeptide with biological activity, where the
CC   polynucleotide hybridises to (I) under stringent hybridisation conditions
CC   or has greater than 99% sequence identity with (I). (I) has respiratory,
CC   cytostatic, antiarthritic, antiinflammatory, gastrointestinal,
CC   antibacterial, immunosuppressive, antidiabetic and antirheumatic
CC   activities, and can be used in gene therapy. (I) can be used for
CC   generating polynucleotides encoding chimeric or fusion proteins and
CC   heterologous protein sequences. The polynucleotides can be used to
CC   express recombinant protein for analysis, characterisation or therapeutic
CC   use; as markers for tissues in which the corresponding protein is
CC   preferentially expressed; as molecular weight markers on gels; as
CC   chromosome markers or tags to identify chromosomes or to map related gene
CC   positions; to compare with endogenous DNA sequences in patients to
CC   identify potential genetic disorders; as probes to hybridise and discover
CC   genes, related DNA sequences; as a source of information to derive PCR
CC   primers for genetic fingerprinting; as a probe to subtract-out known
CC   sequences in the process of discovering other novel polynucleotides; for
CC   selecting and making oligomers for attachment to a gene chip or other
CC   support, including for examination of expression patterns; to raise anti-
CC   protein antibodies using DNA immunisation techniques; and as an antigen
CC   to raise anti-DNA antibodies or elicit another immune response. The
CC   polynucleotides and polypeptides can also be used as nutritional sources
CC   or supplements, e.g. as a protein or amino acid supplement, as a carbon
CC   source, as a nitrogen source or as a source of carbohydrates. The
CC   polynucleotides and polypeptides can also be used treat cancer. The
CC   compositions are useful for promoting better or faster closure of non-
CC   healing wounds, for the generation and regeneration of tissues, for gut
CC   protection or regeneration and treatment of lung or liver fibrosis,
CC   reperfusion injury in various tissues, and conditions resulting from
CC   systemic cytokine damage. The compositions can also be used to treat
CC   inflammatory conditions (e.g. arthritis, inflammatory bowel disease or
CC   Crohn's disease), sepsis, rheumatoid arthritis, diabetes mellitus type 1
CC   or graft versus host disease. The present sequence represents a novel
CC   human polypeptide sequence from the present invention. N.B. The sequences
CC   for this patent were obtained from the USPTO web site from an equivalent
CC   US patent US20040048249A1.
XX
SQ   Sequence 256 AA;

  Query Match             51.4%;  Score 214;  DB 8;  Length 256;
  Best Local Similarity   100.0%;  Pred. No. 3.2e-199;
  Matches  214;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MKRERGALSRASRALRLAPFVYLLLIQTDPLEGVNITSPVRLIHGTVGKSALLSVQYSST 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MKRERGALSRASRALRLAPFVYLLLIQTDPLEGVNITSPVRLIHGTVGKSALLSVQYSST 60

Qy         61 SSDRPVVKWQLKRDKPVTVVQSIGTEVIGTLRPDYRDRIRLFENGSLLLSDLQLADEGTY 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 SSDRPVVKWQLKRDKPVTVVQSIGTEVIGTLRPDYRDRIRLFENGSLLLSDLQLADEGTY 120

Qy        121 EVEISITDDTFTGEKTINLTVDVPISRPQVLVASTTVLELSEAFTLNCSHENGTKPSYTW 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 EVEISITDDTFTGEKTINLTVDVPISRPQVLVASTTVLELSEAFTLNCSHENGTKPSYTW 180

Qy        181 LKDGKPLLNDSRMLLSPDQKVLTITRVLMEDDDL 214
              |||||||||||||||||||||||||||||||||||
Db        181 LKDGKPLLNDSRMLLSPDQKVLTITRVLMEDDDL 214


RESULT 13
ABG75378
ID   ABG75378 standard; protein; 418 AA.
XX
AC   ABG75378;
XX
DT   22-APR-2004  (first entry)
XX
DE   Murine INSP052 complete protein.
XX
```

KW INSP052; human; cell proliferation; autoimmune disease; inflammation;
KW cardiovascular disease; neurological disease; psychiatric disease;
KW developmental disease; metabolic disorder; infection;
KW immunoglobulin domain-containing cell surface recognition molecule.
XX
OS Mus sp.
XX
PN WO2003093316-A2.
XX
PD 13-NOV-2003.
XX
PF 30-APR-2003; 2003WO-GB001851.
XX
PR 30-APR-2002; 2002GB-00009884.
XX
PA (ARES-) ARES TRADING SA.
XX
PI Davids AR, Fagan RJ, Phelps CB, Power C;
XX
DR WPI; 2003-903655/82.
DR N-PSDB; ACH01276.
XX
PT New INSP052 polypeptides and nucleic acids, useful in diagnosing and
PT treating cell proliferative, autoimmune/inflammatory, cardiovascular,
PT neurological, psychiatric, developmental, genetic or metabolic disorder.
XX
PS Example 1; Page 68; 0pp; English.
XX
CC The present invention provides the protein and coding sequences of a
CC novel human immunoglobulin domain-containing cell surface recognition
CC molecule known as INSP052. The polypeptide is useful as immunoglobulin
CC domain-containing cell surface recognition molecule. The sequences may
CC also be used in therapy or diagnosing a disease or in the manufacture of
CC a medicament for treating a disease. The disease is a cell proliferative,
CC autoimmune/inflammatory, cardiovascular, neurological, psychiatric,
CC developmental, genetic or metabolic disorder, an infection or other
CC pathological condition. The polypeptides and nucleic acids are essential
CC to the structural integrity and homeostatic functioning of most tissues.
CC The present sequence is a polypeptide shown in the invention
XX
SQ Sequence 418 AA;

Query Match 36.8%; Score 153; DB 7; Length 418;
Best Local Similarity 100.0%; Pred. No. 1.1e-139;
Matches 153; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 65 PVVKWQLKRDKPVTVVQSIGTEVIGTLRPDYRDRIRLFENGSLLLSDLQLADEGTYEVEI 124
Db 65 PVVKWQLKRDKPVTVVQSIGTEVIGTLRPDYRDRIRLFENGSLLLSDLQLADEGTYEVEI 124

Qy 125 SITDDTFTGEKTINLTVDVPISRPQVLVASTTVLELSEAFTLNCSHENGTKPSYTWLKDG 184
Db 125 SITDDTFTGEKTINLTVDVPISRPQVLVASTTVLELSEAFTLNCSHENGTKPSYTWLKDG 184

Qy 185 KPLLNDSRMLLSPDQKVLTITRVLMEDDDLYSC 217
Db 185 KPLLNDSRMLLSPDQKVLTITRVLMEDDDLYSC 217

RESULT 14
AAM24238
ID AAM24238 standard; protein; 256 AA.
XX
AC AAM24238;
XX
DT 12-OCT-2001 (first entry)
XX
DE Human EST encoded protein SEQ ID NO: 1763.
XX
KW Human; sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;
KW tomato; monkey; dog; sea urchin; expressed sequence tag; EST;
KW diagnostics; forensic test; gene mapping; genetic disorder; biodiversity;
KW gene therapy; nutrition.
XX
OS Homo sapiens.
XX
PN WO200154477-A2.
XX
PD 02-AUG-2001.
XX
PF 25-JAN-2001; 2001WO-US002687.
XX
PR 25-JAN-2000; 2000US-00491404.
PR 17-JUL-2000; 2000US-00617746.
PR 03-AUG-2000; 2000US-00631451.
PR 15-SEP-2000; 2000US-00663870.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V;
PI Cao Y, Drmanac RA, Zhang J, Werhman T;
XX
DR WPI; 2001-476164/51.
DR N-PSDB; AAH98897.
XX
PT Isolated polypeptide for treatment of diseases, diagnostics, raising
PT antibodies and research use.
XX
PS Claim 20; Page 1159-1160; 1275pp; English.
XX
CC The present invention provides the protein and coding sequences of novel
CC proteins from a variety of organisms, including human, dog, cat, horse,
CC cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea
CC urchin and tomato. These were derived from expressed sequence tags (ESTs)
CC from the organism of interest. They can be used in diagnostics,
CC forensics, gene mapping, identification of mutations, to assess
CC biodiversity and for nutritional purposes. The present sequence is a
CC protein of the invention
XX
SQ Sequence 256 AA;

Query Match 36.3%; Score 151; DB 4; Length 256;
Best Local Similarity 100.0%; Pred. No. 6.2e-138;
Matches 151; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MKRERGALSRASRALRLAPFVYLLLIQTDPLEGVNITSPVRLIHGTVGKSALLSVQYSST 60
Db 1 MKRERGALSRASRALRLAPFVYLLLIQTDPLEGVNITSPVRLIHGTVGKSALLSVQYSST 60

Qy 61 SSDRPVVKWQLKRDKPVTVVQSIGTEVIGTLRPDYRDRIRLFENGSLLLSDLQLADEGTY 120
Db 61 SSDRPVVKWQLKRDKPVTVVQSIGTEVIGTLRPDYRDRIRLFENGSLLLSDLQLADEGTY 120

Qy 121 EVEISITDDTFTGEKTINLTVDVPISRPQVL 151
Db 121 EVEISITDDTFTGEKTINLTVDVPISRPQVL 151

RESULT 15
ADM87787
ID ADM87787 standard; protein; 256 AA.
XX
AC ADM87787;
XX
DT 03-JUN-2004 (first entry)
XX
DE Human EST derived amino acid sequence SEQ ID NO:880.
XX
KW respiratory; cytostatic; antiarthritic; antiinflammatory;
KW gastrointestinal; antibacterial; immunosuppressive; antidiabetic;
KW antirheumatic; gene therapy; molecular weight marker; chromosome marker;
KW chromosome tag; genetic fingerprinting; nutritional supplement; cancer;
KW inflammatory condition; arthritis; inflammatory bowel disease;
KW Crohn's disease; sepsis; rheumatoid arthritis; diabetes mellitus type 1;
KW graft versus host disease; human; expressed sequence tag; EST.
XX

OS Homo sapiens.
XX
PN WO2004009834-A2.
XX
PD 29-JAN-2004.
XX
PF 19-JUL-2002; 2002WO-US022858.
XX
PR 21-JUL-2001; 2001US-0306971P.
PR 28-MAR-2002; 2002US-00112944.
XX
PA (NUVE-) NUVELO INC.
XX
PI Tang YT, Yang Y, Weng G, Zhang J, Ren F, Xue A, Wang J;
PI Wehrman T, Ghosh MJ, Wang D, Zhao QA, Wang Z;
XX
DR WPI; 2004-143291/14.
DR N-PSDB; ADM87569.
XX
PT New isolated polynucleotides and polypeptides, useful for treating, e.g.
PT cancer, lung or liver fibrosis, arthritis, inflammatory bowel disease,
PT Crohn's disease, rheumatoid arthritis, diabetes mellitus type 1 or graft
PT versus host disease.
XX
PS Example 2; SEQ ID NO 880; 591pp; English.
XX
CC The present invention describes an isolated polynucleotide (I): (a)
CC comprising a nucleotide sequence selected from SEQ ID NO:1-244; or (b)
CC which encodes a polypeptide with biological activity, where the
CC polynucleotide hybridises to (I) under stringent hybridisation conditions
CC or has greater than 99% sequence identity with (I). (I) has respiratory,
CC cytostatic, antiarthritic, antiinflammatory, gastrointestinal,
CC antibacterial, immunosuppressive, antidiabetic and antirheumatic
CC activities, and can be used in gene therapy. (I) can be used for
CC generating polynucleotides encoding chimeric or fusion proteins and
CC heterologous protein sequences. The polynucleotides can be used to
CC express recombinant protein for analysis, characterisation or therapeutic
CC use; as markers for tissues in which the corresponding protein is
CC preferentially expressed; as molecular weight markers on gels; as
CC chromosome markers or tags to identify chromosomes or to map related gene
CC positions; to compare with endogenous DNA sequences in patients to
CC identify potential genetic disorders; as probes to hybridise and discover
CC genes, related DNA sequences; as a source of information to derive PCR
CC primers for genetic fingerprinting; as a probe to subtract-out known
CC sequences in the process of discovering other novel polynucleotides; for
CC selecting and making oligomers for attachment to a gene chip or other
CC support, including for examination of expression patterns; to raise anti-
CC protein antibodies using DNA immunisation techniques; and as an antigen
CC to raise anti-DNA antibodies or elicit another immune response. The
CC polynucleotides and polypeptides can also be used as nutritional sources
CC or supplements, e.g. as a protein or amino acid supplement, as a carbon
CC source, as a nitrogen source or as a source of carbohydrates. The
CC polynucleotides and polypeptides can also be used treat cancer. The
CC compositions are useful for promoting better or faster closure of non-
CC healing wounds, for the generation and regeneration of tissues, for gut
CC protection or regeneration and treatment of lung or liver fibrosis,
CC reperfusion injury in various tissues, and conditions resulting from
CC systemic cytokine damage. The compositions can also be used to treat
CC inflammatory conditions (e.g. arthritis, inflammatory bowel disease or
CC Crohn's disease), sepsis, rheumatoid arthritis, diabetes mellitus type 1
CC or graft versus host disease. The present sequence represents an
CC expressed sequence tag (EST) derived amino acid sequence from the present
CC invention. N.B. The sequences for this patent were obtained from the
CC USPTO web site from an equivalent US patent US20040048249A1.
XX
SQ Sequence 256 AA;

Query Match 36.3%; Score 151; DB 8; Length 256;
Best Local Similarity 100.0%; Pred. No. 6.2e-138;
Matches 151; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MKRERGALSRASRALRLAPFVYLLLIQTDPLEGVNITSPVRLIHGTVGKSALLSVQYSST 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MKRERGALSRASRALRLAPFVYLLLIQTDPLEGVNITSPVRLIHGTVGKSALLSVQYSST 60

Qy         61 SSDRPVVKWQLKRDKPVTVVQSIGTEVIGTLRPDYRDRIRLFENGSLLLSDLQLADEGTY 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 SSDRPVVKWQLKRDKPVTVVQSIGTEVIGTLRPDYRDRIRLFENGSLLLSDLQLADEGTY 120

Qy        121 EVEISITDDTFTGEKTINLTVDVPISRPQVL 151
              |||||||||||||||||||||||||||||||
Db        121 EVEISITDDTFTGEKTINLTVDVPISRPQVL 151
```

Search completed: August 3, 2005, 10:05:37
Job time : 172 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: August 3, 2005, 10:02:38 ; Search time 169 Seconds
(without alignments)
1260.503 Million cell updates/sec

Title: US-10-706-691-16
Perfect score: 416
Sequence: 1 MKRERGALSRASRALRLAPF.....TAGVHIIREQDEAGPVEISA 416

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 1612378 seqs, 512079187 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : UniProt_03:*
1: uniprot_sprot:*
2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	315	75.7	416	2	Q8N7I3	Q8n7i3 homo sapien
2	313	75.2	416	2	Q67IP8	Q67ip8 homo sapien
3	153	36.8	413	2	Q640R3	Q640r3 mus musculu
4	123	29.6	165	2	Q8ND35	Q8nd35 homo sapien
5	122	29.3	367	2	Q6ZWL4	Q6zwl4 homo sapien
6	9	2.2	691	2	Q8A099	Q8a099 bacteroides
7	8	1.9	183	2	Q8SAV3	Q8sav3 oryza sativ
8	8	1.9	183	2	Q7XGC4	Q7xgc4 oryza sativ
9	8	1.9	226	2	Q7RA64	Q7ra64 plasmodium
10	8	1.9	251	2	Q8LMI8	Q8lmi8 oryza sativ
11	8	1.9	253	2	Q97JP1	Q97jp1 clostridium
12	8	1.9	263	2	Q6L9V2	Q6l9v2 neocalanus
13	8	1.9	263	2	Q7NNR0	Q7nnr0 gloeobacter
14	8	1.9	287	2	Q6I6X5	Q6i6x5 oryzias lat
15	8	1.9	288	2	Q8JIN6	Q8jin6 oryzias lat
16	8	1.9	297	2	Q6I6X2	Q6i6x2 oryzias lat
17	8	1.9	297	2	Q6I6X7	Q6i6x7 oryzias lat
18	8	1.9	303	2	Q89F66	Q89f66 bradyrhizob
19	8	1.9	305	1	LPXC_ECOLI	P07652 escherichia
20	8	1.9	305	1	LPXC_SALTI	Q8z9g5 salmonella
21	8	1.9	305	1	LPXC_SALTY	Q8zrt9 salmonella
22	8	1.9	306	2	Q6LHY4	Q6lhy4 photobacter
23	8	1.9	321	2	Q6I6X6	Q6i6x6 oryzias lat
24	8	1.9	335	2	Q7U9L4	Q7u9l4 synechococc
25	8	1.9	337	2	Q8S716	Q8s716 oryza sativ
26	8	1.9	340	1	IGB1_RAT	O08836 rattus norv
27	8	1.9	345	2	Q7V4S9	Q7v4s9 prochloroco
28	8	1.9	350	2	Q03130	Q03130 saccharomyc
29	8	1.9	479	2	Q9ULG2	Q9ulg2 homo sapien
30	8	1.9	480	1	MTH4_DROME	Q9v817 drosophila
31	8	1.9	500	2	Q6XL69	Q6xl69 rutilus rut
32	8	1.9	501	2	Q8AV31	Q8av31 brachydanio
33	8	1.9	511	1	MTH3_DROME	Q9v818 drosophila
34	8	1.9	514	2	Q6NNZ3	Q6nnz3 drosophila
35	8	1.9	517	2	Q86BE7	Q86be7 drosophila
36	8	1.9	577	2	Q6FV98	Q6fv98 candida gla
37	8	1.9	599	2	Q8KQY2	Q8kqy2 vibrio chol
38	8	1.9	606	2	Q8RL14	Q8rl14 providencia
39	8	1.9	619	2	Q08548	Q08548 saccharomyc
40	8	1.9	666	2	Q8VW62	Q8vw62 bacillus th
41	8	1.9	744	1	HXC1_HAEIN	P44523 haemophilus
42	8	1.9	841	2	Q7TS80	Q7tss0 mus musculu
43	8	1.9	862	2	Q6C455	Q6c455 yarrowia li
44	8	1.9	1028	2	Q6INB5	Q6inb5 xenopus lae
45	8	1.9	1184	2	Q66K08	Q66k08 mus musculu

ALIGNMENTS

RESULT 1
Q8N7I3
ID Q8N7I3 PRELIMINARY; PRT; 416 AA.
AC Q8N7I3;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Hypothetical protein FLJ25530.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RA Tashiro H., Yamazaki M., Watanabe K., Kumagai A., Itakura S.,
RA Fukuzumi Y., Fujimori Y., Komiyama M., Suzuki Y., Hata H.,
RA Nakagawa K., Mizuno S., Morinaga M., Kawamura M., Sugiyama T.,
RA Irie R., Otsuki T., Sato H., Nishikawa T., Sugiyama A., Kawakami B.,
RA Nagai K., Isogai T., Sugano S.;
RL Submitted (JUL-2002) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AK098396; BAC05297.1; -.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003598; Ig_c2.
DR Pfam; PF00047; ig; 1.
DR SMART; SM00408; IGc2; 1.
DR PROSITE; PS50835; IG_LIKE; 1.
SQ SEQUENCE 416 AA; 45994 MW; 47120CA9A00EE1CF CRC64;

Query Match 75.7%; Score 315; DB 2; Length 416;
Best Local Similarity 99.8%; Pred. No. 2.3e-302;
Matches 415; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy          1 MKRERGALSRASRALRLAPFVYLLLIQTDPLEGVNITSPVRLIHGTVGKSALLSVQYSST 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MKRERGALSRASRALRLAPFVYLLLIQTDPLEGVNITSPVRLIHGTVGKSALLSVQYSST 60

Qy         61 SSDRPVVKWQLKRDKPVTVVQSIGTEVIGTLRPDYRDRIRLFENGSLLLSDLQLADEGTY 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 SSDRPVVKWQLKRDKPVTVVQSIGTEVIGTLRPDYRDRIRLFENGSLLLSDLQLADEGTY 120

Qy        121 EVEISITDDTFTGEKTINLTVDVPISRPQVLVASTTVLELSEAFTLNCSHENGTKPSYTW 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 EVEISITDDTFTGEKTINLTVDVPISRPQVLVASTTVLELSEAFTLNCSHENGTKPSYTW 180

Qy        181 LKDGKPLLNDSRMLLSPDQKVLTITRVLMEDDDLYSCMVENPISQGRSLPVKITVYRRSS 240
              |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db        181 LKDGKPLLNDSRMLLSPDQKVLTITRVLMEDDDLYSCVVENPISQGRSLPVKITVYRRSS 240

Qy        241 LYIILSTGGIFLLVTLVTVCACWKPSKRKQKKLEKQNSLEYMDQNDDRLKPEADTLPRSG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 LYIILSTGGIFLLVTLVTVCACWKPSKRKQKKLEKQNSLEYMDQNDDRLKPEADTLPRSG 300

Qy        301 EQERKNPMALYILKDKDSPETEENPAPEPRSATEPGPPGYSVSPAVPGRSPGLPIRSARR 360
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 EQERKNPMALYILKDKDSPETEENPAPEPRSATEPGPPGYSVSPAVPGRSPGLPIRSARR 360

Qy        361 YPRSPARSPATGRTHSSPPRAPSSPGRSRSASRTLRTAGVHIIREQDEAGPVEISA 416
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 YPRSPARSPATGRTHSSPPRAPSSPGRSRSASRTLRTAGVHIIREQDEAGPVEISA 416


RESULT 2
Q67IP8
ID   Q67IP8      PRELIMINARY;      PRT;   416 AA.
AC   Q67IP8;
DT   25-OCT-2004 (TrEMBLrel. 28, Created)
DT   25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT   25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE   Hypothetical protein.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RA   Shen S., Moh M.C.;
RT   "A gene related to human hepatocellular carcinoma.";
RL   Submitted (JUL-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY047587; AAQ93018.1; -.
DR   InterPro; IPR003599; Ig.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003598; Ig_c2.
DR   Pfam; PF00047; ig; 1.
DR   SMART; SM00409; IG; 2.
DR   SMART; SM00408; IGc2; 1.
DR   PROSITE; PS50835; IG_LIKE; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   416 AA;  46055 MW;  7B8882298BEB4ABF CRC64;

  Query Match             75.2%;  Score 313;  DB 2;  Length 416;
  Best Local Similarity   99.8%;  Pred. No. 2.2e-300;
  Matches  413;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          3 RERGALSRASRALRLAPFVYLLLIQTDPLEGVNITSPVRLIHGTVGKSALLSVQYSSTSS 62
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          3 RERGALSRASRALRLAPFVYLLLIQTDPLEGVNITSPVRLIHGTVGKSALLSVQYSSTSS 62

Qy         63 DRPVVKWQLKRDKPVTVVQSIGTEVIGTLRPDYRDRIRLFENGSLLLSDLQLADEGTYEV 122
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         63 DRPVVKWQLKRDKPVTVVQSIGTEVIGTLRPDYRDRIRLFENGSLLLSDLQLADEGTYEV 122

Qy        123 EISITDDTFTGEKTINLTVDVPISRPQVLVASTTVLELSEAFTLNCSHENGTKPSYTWLK 182
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        123 EISITDDTFTGEKTINLTVDVPISRPQVLVASTTVLELSEAFTLNCSHENGTKPSYTWLK 182

Qy        183 DGKPLLNDSRMLLSPDQKVLTITRVLMEDDDLYSCMVENPISQGRSLPVKITVYRRSSLY 242
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        183 DGKPLLNDSRMLLSPDQKVLTITRVLMEDDDLYSCMVENPISQGRSLPVKITVYRRSSLY 242

Qy        243 IILSTGGIFLLVTLVTVCACWKPSKRKQKKLEKQNSLEYMDQNDDRLKPEADTLPRSGEQ 302
                |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db        243 IILSTGGIFLLVTLVTVCACWKPSKRKQKKLEKQNSLEYMDRNDDRLKPEADTLPRSGEQ 302

Qy        303 ERKNPMALYILKDKDSPETEENPAPEPRSATEPGPPGYSVSPAVPGRSPGLPIRSARRYP 362
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        303 ERKNPMALYILKDKDSPETEENPAPEPRSATEPGPPGYSVSPAVPGRSPGLPIRSARRYP 362

Qy        363 RSPARSPATGRTHSSPPRAPSSPGRSRSASRTLRTAGVHIIREQDEAGPVEISA 416
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        363 RSPARSPATGRTHSSPPRAPSSPGRSRSASRTLRTAGVHIIREQDEAGPVEISA 416


RESULT 3
Q640R3
ID   Q640R3      PRELIMINARY;      PRT;   413 AA.
AC   Q640R3;
DT   25-OCT-2004 (TrEMBLrel. 28, Created)
DT   25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT   25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE   2900042E01Rik protein (Fragment).
GN   Name=2900042E01Rik;
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6; TISSUE=Brain;
RX   PubMed=12477932; DOI=10.1073/pnas.242603899;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA   Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA   Jones S.J., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length human
RT   and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6; TISSUE=Brain;
RA   Director MGC Project;
RL   Submitted (SEP-2004) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC082537; AAH82537.1; -.
FT   NON_TER       1      1
SQ   SEQUENCE   413 AA;  45665 MW;  B6EFCA2D6D2CA3C1 CRC64;

  Query Match             36.8%;  Score 153;  DB 2;  Length 413;
  Best Local Similarity   100.0%;  Pred. No. 3.9e-142;
  Matches  153;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         65 PVVKWQLKRDKPVTVVQSIGTEVIGTLRPDYRDRIRLFENGSLLLSDLQLADEGTYEVEI 124
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         60 PVVKWQLKRDKPVTVVQSIGTEVIGTLRPDYRDRIRLFENGSLLLSDLQLADEGTYEVEI 119

Qy        125 SITDDTFTGEKTINLTVDVPISRPQVLVASTTVLELSEAFTLNCSHENGTKPSYTWLKDG 184
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        120 SITDDTFTGEKTINLTVDVPISRPQVLVASTTVLELSEAFTLNCSHENGTKPSYTWLKDG 179

Qy        185 KPLLNDSRMLLSPDQKVLTITRVLMEDDDLYSC 217
                |||||||||||||||||||||||||||||||
Db        180 KPLLNDSRMLLSPDQKVLTITRVLMEDDDLYSC 212


RESULT 4
Q8ND35
ID   Q8ND35      PRELIMINARY;      PRT;   165 AA.
AC   Q8ND35;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Hypothetical protein DKFZp547O159 (Fragment).
GN   Name=DKFZp547O159;
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
```

```
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RA   Bloecker H., Boecher M., Brandt P., Mewes H.W., Weil B., Wiemann S.;
RL   Submitted (JUL-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL834419; CAD39081.1; -.
KW   Hypothetical protein.
FT   NON_TER       1      1
SQ   SEQUENCE   165 AA;  18161 MW;  5052FA978C437486 CRC64;

  Query Match             29.6%;  Score 123;  DB 2;  Length 165;
  Best Local Similarity   100.0%;  Pred. No. 7.8e-113;
  Matches  123;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        294 DTLPRSGEQERKNPMALYILKDKDSPETEENPAPEPRSATEPGPPGYSVSPAVPGRSPGL 353
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         43 DTLPRSGEQERKNPMALYILKDKDSPETEENPAPEPRSATEPGPPGYSVSPAVPGRSPGL 102

Qy        354 PIRSARRYPRSPARSPATGRTHSSPPRAPSSPGRSRSASRTLRTAGVHIIREQDEAGPVE 413
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        103 PIRSARRYPRSPARSPATGRTHSSPPRAPSSPGRSRSASRTLRTAGVHIIREQDEAGPVE 162

Qy        414 ISA 416
              |||
Db        163 ISA 165


RESULT 5
Q6ZWL4
ID   Q6ZWL4      PRELIMINARY;      PRT;   367 AA.
AC   Q6ZWL4;
DT   05-JUL-2004 (TrEMBLrel. 27, Created)
DT   05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT   05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE   Hypothetical protein FLJ16002.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RA   Suzuki O., Sasaki N., Aotsuka S., Shoji T., Ichihara T., Shiohata N.,
RA   Matsumoto K., Hirano M., Sano S., Nomura R., Yoshikawa Y.,
RA   Matsumura Y., Moriya S., Chiba E., Momiyama H., Onogawa S.,
RA   Kaeriyama S., Satoh N., Matsunawa H., Takahashi E., Kataoka R.,
RA   Kuga N., Kuroda A., Satoh I., Kamata K., Takami S., Terashima Y.,
RA   Watanabe M., Sugiyama T., Irie R., Otsuki T., Sato H., Ota T.,
RA   Wakamatsu A., Ishii S., Yamamoto J., Isono Y., Kawai-Hio Y., Saito K.,
RA   Nishikawa T., Kimura K., Yamashita H., Matsuo K., Nakamura Y.,
RA   Sekine M., Kikuchi H., Kanda K., Wagatsuma M., Murakawa K.,
RA   Kanehori K., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B.,
RA   Suzuki Y., Sugano S., Nagahari K., Masuho Y., Nagai K., Isogai T.;
RL   Submitted (JUL-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AK122595; BAC85486.1; -.
DR   GO; GO:0004872; F:receptor activity; IEA.
DR   InterPro; IPR003599; Ig.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003598; Ig_c2.
DR   Pfam; PF00047; ig; 1.
DR   SMART; SM00409; IG; 2.
DR   SMART; SM00408; IGc2; 1.
DR   PROSITE; PS50835; IG_LIKE; 1.
KW   Receptor.
SQ   SEQUENCE   367 AA;  40456 MW;  35956FA245A408F0 CRC64;

  Query Match             29.3%;  Score 122;  DB 2;  Length 367;
  Best Local Similarity   100.0%;  Pred. No. 1.6e-111;
  Matches  122;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         96 RDRIRLFENGSLLLSDLQLADEGTYEVEISITDDTFTGEKTINLTVDVPISRPQVLVAST 155
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         96 RDRIRLFENGSLLLSDLQLADEGTYEVEISITDDTFTGEKTINLTVDVPISRPQVLVAST 155

Qy        156 TVLELSEAFTLNCSHENGTKPSYTWLKDGKPLLNDSRMLLSPDQKVLTITRVLMEDDDLY 215
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        156 TVLELSEAFTLNCSHENGTKPSYTWLKDGKPLLNDSRMLLSPDQKVLTITRVLMEDDDLY 215

Qy        216 SC 217
              ||
Db        216 SC 217


RESULT 6
Q8A099
ID   Q8A099      PRELIMINARY;      PRT;   691 AA.
AC   Q8A099;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Putative outer membrane protein, probably involved in nutrient
DE   binding.
GN   OrderedLocusNames=BT4122;
OS   Bacteroides thetaiotaomicron.
OC   Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC   Bacteroidaceae; Bacteroides.
OX   NCBI_TaxID=818;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=VPI-5482 / ATCC 29148;
RX   MEDLINE=22550858; PubMed=12663928; DOI=10.1126/science.1080029;
RA   Xu J., Bjursell M.K., Himrod J., Deng S., Carmichael L.K.,
RA   Chiang H.C., Hooper L.V., Gordon J.I.;
RT   "A genomic view of the human-Bacteroides thetaiotaomicron symbiosis.";
RL   Science 299:2074-2076(2003).
DR   EMBL; AE016943; AAO79227.1; -.
KW   Complete proteome.
SQ   SEQUENCE   691 AA;  77418 MW;  A99BAC4FD2C6667C CRC64;

  Query Match              2.2%;  Score 9;  DB 2;  Length 691;
  Best Local Similarity   100.0%;  Pred. No. 16;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         32 EGVNITSPV 40
              |||||||||
Db        428 EGVNITSPV 436


RESULT 7
Q8SAV3
ID   Q8SAV3      PRELIMINARY;      PRT;   183 AA.
AC   Q8SAV3;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hypothetical protein OSJNBa0051J07.10.
GN   Name=OSJNBa0051J07.10;
OS   Oryza sativa (Rice).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC   Ehrhartoideae; Oryzeae; Oryza.
OX   NCBI_TaxID=4530;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Wing R.A., Yu Y., Soderlund C., Chen M., Kim H.-R., Rambo T.,
RA   Saski C., Henry D., Oates R., Simmons J.;
RL   Submitted (FEB-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC098566; AAL77123.1; -.
DR   Gramene; Q8SAV3; -.
KW   Hypothetical protein.
SQ   SEQUENCE   183 AA;  20011 MW;  37F2293DCA9196AC CRC64;

  Query Match              1.9%;  Score 8;  DB 2;  Length 183;
  Best Local Similarity   100.0%;  Pred. No. 47;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy        387 RSRSASRT 394
              ||||||||
Db         76 RSRSASRT 83


RESULT 8
Q7XGC4
ID   Q7XGC4      PRELIMINARY;      PRT;   183 AA.
AC   Q7XGC4;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hypothetical protein.
GN   ORFNames=OSJNBa0051J07.10;
OS   Oryza sativa (japonica cultivar-group).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC   Ehrhartoideae; Oryzeae; Oryza.
OX   NCBI_TaxID=39947;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   The Rice Chromosome 10 Sequencing Consortium;
RT   "In-depth view of structure, activity, and evolution of rice
RT   chromosome 10.";
RL   Science 300:1566-1569(2003).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Buell C.R., Wing R.A., McCombie W.R., Messing J., Yuan Q.;
RL   Submitted (MAY-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AE017064; AAP52407.1; -.
DR   Gramene; Q7XGC4; -.
KW   Hypothetical protein.
SQ   SEQUENCE   183 AA;  20011 MW;  37F2293DCA9196AC CRC64;

  Query Match             1.9%;  Score 8;  DB 2;  Length 183;
  Best Local Similarity   100.0%;  Pred. No. 47;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        387 RSRSASRT 394
              ||||||||
Db         76 RSRSASRT 83


RESULT 9
Q7RA64
ID   Q7RA64      PRELIMINARY;      PRT;   226 AA.
AC   Q7RA64;
DT   01-MAR-2004 (TrEMBLrel. 26, Created)
DT   01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Hypothetical protein.
GN   Name=PY06639;
OS   Plasmodium yoelii yoelii.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=73239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=17XNL;
RX   PubMed=12368865; DOI=10.1038/nature01099;
RA   Carlton J.M., Angiuoli S.V., Suh B.B., Kooij T.W., Pertea M.,
RA   Silva J.C., Ermolaeva M.D., Allen J.E., Selengut J.D., Koo H.L.,
RA   Peterson J.D., Pop M., Kosack D.S., Shumway M.F., Bidwell S.L.,
RA   Shallom S.J., van Aken S.E., Riedmuller S.B., Feldblyum T.V.,
RA   Cho J.K., Quackenbush J., Sedegah M., Shoaibi A., Cummings L.M.,
RA   Florens L., Yates F.R. III, Raine J.D., Sinden R.E., Harris M.A.,
RA   Cunningham D.A., Preiser P.R., Bergman L.W., Vaidya A.B.,
RA   van Lin L.H., Janse C.J., Waters A.P., Smith H.O., White O.R.,
RA   Salzberg S.L., Venter J.C., Fraser C.M., Hoffman S.L., Gardner M.J.,
RA   Carucci D.J.;
RT   "Genome sequence and comparative analysis of the model rodent malaria
RT   parasite Plasmodium yoelii yoelii.";
RL   Nature 419:512-519(2002).
CC   -!- CAUTION: The sequence shown here is derived from an
CC       EMBL/GenBank/DDBJ whole genome shotgun (WGS) entry which is
CC       preliminary data.
DR   EMBL; AABL01002273; EAA18887.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   226 AA;  26674 MW;  CA5A0F07D1496B3A CRC64;

  Query Match             1.9%;  Score 8;  DB 2;  Length 226;
  Best Local Similarity   100.0%;  Pred. No. 57;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        269 KQKKLEKQ 276
              ||||||||
Db         35 KQKKLEKQ 42


RESULT 10
Q8LMI8
ID   Q8LMI8      PRELIMINARY;      PRT;   251 AA.
AC   Q8LMI8;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Hypothetical protein OSJNBb0038A07.26.
GN   Name=OSJNBb0038A07.26;
OS   Oryza sativa (japonica cultivar-group).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC   Ehrhartoideae; Oryzeae; Oryza.
OX   NCBI_TaxID=39947;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Buell C.R., Yuan Q., Ouyang S., Liu J., Gansberger K., Kim M.M.,
RA   Overton II L.L., Bera J.J., Tsitrin T., Krol M.I., Jarrahi B.B.,
RA   Jin S.S., Koo H., Zismann V., Hsiao J., Blunt S., Vanaken S.S.,
RA   Utterback T.T., Feldblyum T.V., Yang Q.Q., Haas B.J., Suh B.B.,
RA   Peterson J.J., Quackenbush J., White O., Salzberg S.L., Fraser C.M.;
RL   Submitted (MAR-2002) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Buell R.;
RL   Submitted (JUN-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC113948; AAM94543.1; -.
DR   Gramene; Q8LMI8; -.
KW   Hypothetical protein.
SQ   SEQUENCE   251 AA;  27363 MW;  CA05BAF6DF0927C7 CRC64;

  Query Match             1.9%;  Score 8;  DB 2;  Length 251;
  Best Local Similarity   100.0%;  Pred. No. 63;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 RERGALSR 10
              ||||||||
Db        196 RERGALSR 203


RESULT 11
Q97JP1
ID   Q97JP1      PRELIMINARY;      PRT;   253 AA.
AC   Q97JP1;
DT   01-OCT-2001 (TrEMBLrel. 18, Created)
DT   01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Predicted lytic murein transglycosylase (N-term. LysM motif repeat
DE   domain).
GN   OrderedLocusNames=CAC1232;
OS   Clostridium acetobutylicum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1488;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 824 / DSM 792 / VKM B-1787;
RX   MEDLINE=21359325; PubMed=11466286;
```

```
RX   DOI=10.1128/JB.183.16.4823-4838.2001;
RA   Noelling J., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q.,
RA   Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,
RA   Tatusov R.L., Sabathe F., Doucette-Stamm L.A., Soucaille P.,
RA   Daly M.J., Bennett G.N., Koonin E.V., Smith D.R.;
RT   "Genome sequence and comparative analysis of the solvent-producing
RT   bacterium Clostridium acetobutylicum.";
RL   J. Bacteriol. 183:4823-4838(2001).
DR   EMBL; AE007636; AAK79204.1; -.
DR   PIR; A97052; A97052.
DR   GO; GO:0016998; P:cell wall catabolism; IEA.
DR   InterPro; IPR002482; LysM.
DR   InterPro; IPR011105; SleB_hydro.
DR   Pfam; PF07486; Hydrolase_2; 1.
DR   Pfam; PF01476; LysM; 2.
DR   SMART; SM00257; LysM; 2.
KW   Complete proteome.
SQ   SEQUENCE   253 AA;  26927 MW;  5B1A03BD48BB09C3 CRC64;

  Query Match             1.9%;  Score 8;  DB 2;  Length 253;
  Best Local Similarity   100.0%;  Pred. No. 64;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        108 LLSDLQLA 115
              ||||||||
Db         94 LLSDLQLA 101


RESULT 12
Q6L9V2
ID   Q6L9V2      PRELIMINARY;      PRT;   263 AA.
AC   Q6L9V2;
DT   05-JUL-2004 (TrEMBLrel. 27, Created)
DT   05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT   05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE   Cytochrome c oxidase subunits III.
GN   Name=COIII;
OS   Neocalanus cristatus.
OG   Mitochondrion.
OC   Eukaryota; Metazoa; Arthropoda; Crustacea; Maxillopoda; Copepoda;
OC   Calanoida; Calanidae; Neocalanus.
OX   NCBI_TaxID=119368;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   PubMed=15145056;
RA   Machida R.J., Miya M.U., Nishida M., Nishida S.;
RT   "Large-scale gene rearrangements in the mitochondrial genomes of two
RT   calanoid copepods Eucalanus bungii and Neocalanus cristatus
RT   (Crustacea), with notes on new versatile primers for the srRNA and COI
RT   genes.";
RL   Gene 332:71-78(2004).
CC   -!- FUNCTION: Subunits I, II and III form the functional core of the
CC       enzyme complex (By similarity).
CC   -!- CATALYTIC ACTIVITY: 4 ferrocytochrome c + O(2) = 4 ferricytochrome
CC       c + 2 H(2)O.
CC   -!- SIMILARITY: Belongs to the cytochrome c oxidase subunit 3 family.
DR   EMBL; AB091773; BAD19006.1; -.
DR   GO; GO:0016021; C:integral to membrane; IEA.
DR   GO; GO:0005739; C:mitochondrion; IEA.
DR   GO; GO:0004129; F:cytochrome-c oxidase activity; IEA.
DR   GO; GO:0016491; F:oxidoreductase activity; IEA.
DR   GO; GO:0006118; P:electron transport; IEA.
DR   InterPro; IPR000298; CytC_oxdse_III.
DR   Pfam; PF00510; COX3; 1.
DR   ProDom; PD000382; CytC_oxdse_III; 1.
DR   PROSITE; PS50253; COX3; 1.
KW   Mitochondrion; Oxidoreductase; Transmembrane.
SQ   SEQUENCE   263 AA;  30086 MW;  F277429A637D861F CRC64;

  Query Match             1.9%;  Score 8;  DB 2;  Length 263;
  Best Local Similarity   100.0%;  Pred. No. 66;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        250 IFLLVTLV 257
              ||||||||
Db        214 IFLLVTLV 221


RESULT 13
Q7NNR0
ID   Q7NNR0      PRELIMINARY;      PRT;   263 AA.
AC   Q7NNR0;
DT   01-MAR-2004 (TrEMBLrel. 26, Created)
DT   01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Gll0350 protein.
GN   OrderedLocusNames=gll0350;
OS   Gloeobacter violaceus.
OC   Bacteria; Cyanobacteria; Chroococcales; Gloeobacter.
OX   NCBI_TaxID=33072;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=PCC 7421;
RX   MEDLINE=22977040; PubMed=14621292;
RA   Nakamura Y., Kaneko T., Sato S., Mimuro M., Miyashita H., Tsuchiya T.,
RA   Sasamoto S., Watanabe A., Kawashima K., Kishida Y., Kiyokawa C.,
RA   Kohara M., Matsumoto M., Matsuno A., Nakazaki N., Shimpo S.,
RA   Takeuchi C., Yamada M., Tabata S.;
RT   "Complete genome structure of Gloeobacter violaceus PCC 7421, a
RT   cyanobacterium that lacks thylakoids.";
RL   DNA Res. 10:137-145(2003).
DR   EMBL; AP006569; BAC88291.1; -.
KW   Complete proteome.
SQ   SEQUENCE   263 AA;  30782 MW;  5B05B2B2007DE17F CRC64;

  Query Match             1.9%;  Score 8;  DB 2;  Length 263;
  Best Local Similarity   100.0%;  Pred. No. 66;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        351 PGLPIRSA 358
              ||||||||
Db          6 PGLPIRSA 13


RESULT 14
Q6I6X5
ID   Q6I6X5      PRELIMINARY;      PRT;   287 AA.
AC   Q6I6X5;
DT   05-JUL-2004 (TrEMBLrel. 27, Created)
DT   05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT   05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE   Transformer-2b4.
GN   Name=Tra2b4;
OS   Oryzias latipes (Medaka fish) (Japanese ricefish).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC   Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
OC   Beloniformes; Adrianichthyidae; Oryziinae; Oryzias.
OX   NCBI_TaxID=8090;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Shiraishi E., Imazato H., Yamamoto T., Yokoi H., Abe S., Kitano T.;
RT   "Identification of two teleost homologs of the Drosophila sex
RT   determination factor, transformer-2 in medaka (Oryzias latipes).";
RL   Mech. Dev. 121:991-996(2004).
DR   EMBL; AB159273; BAD24703.1; -.
DR   InterPro; IPR000504; RNA_rec_mot.
DR   Pfam; PF00076; RRM_1; 1.
DR   SMART; SM00360; RRM; 1.
DR   PROSITE; PS50102; RRM; 1.
DR   PROSITE; PS00030; RRM_RNP_1; UNKNOWN_1.
SQ   SEQUENCE   287 AA;  33003 MW;  ECBA9D3743027E2B CRC64;

  Query Match             1.9%;  Score 8;  DB 2;  Length 287;
  Best Local Similarity   100.0%;  Pred. No. 71;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy        363 RSPARSPA 370
              ||||||||
Db         27 RSPARSPA 34


RESULT 15
Q8JIN6
ID   Q8JIN6      PRELIMINARY;      PRT;   288 AA.
AC   Q8JIN6;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Transformer-2b.
GN   Name=Tra2b;
OS   Oryzias latipes (Medaka fish) (Japanese ricefish).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC   Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
OC   Beloniformes; Adrianichthyidae; Oryziinae; Oryzias.
OX   NCBI_TaxID=8090;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Shiraishi E., Imazato H., Yamamoto T., Yokoi H., Abe S., Kitano T.;
RT   "Identification of two teleost homologs of the Drosophila sex
RT   determination factor, transformer-2 in medaka (Oryzias latipes).";
RL   Mech. Dev. 121:991-996(2004).
DR   EMBL; AB079122; BAC06514.1; -.
DR   HSSP; Q9V535; 1HL6.
DR   InterPro; IPR000504; RNA_rec_mot.
DR   Pfam; PF00076; RRM_1; 1.
DR   SMART; SM00360; RRM; 1.
DR   PROSITE; PS50102; RRM; 1.
DR   PROSITE; PS00030; RRM_RNP_1; UNKNOWN_1.
SQ   SEQUENCE   288 AA;  32991 MW;  8B92AA9D1326FFF2 CRC64;

  Query Match             1.9%;  Score 8;  DB 2;  Length 288;
  Best Local Similarity   100.0%;  Pred. No. 72;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        363 RSPARSPA 370
              ||||||||
Db         28 RSPARSPA 35


Search completed: August  3, 2005, 10:08:31
Job time : 171 secs
```

OM protein - protein search, using sw model

Run on: August 3, 2005, 10:02:38 ; Search time 41 Seconds
(without alignments)
976.248 Million cell updates/sec

Title: US-10-706-691-16
Perfect score: 416
Sequence: 1 MKRERGALSRASRALRLAPF.....TAGVHIIREQDEAGPVEISA 416

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 283416 seqs, 96216763 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : PIR_79:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	8	1.9	253	2	A97052	probable lytic mur
2	8	1.9	305	1	BVECEA	UDP-3-O-[3-hydroxy
3	8	1.9	305	2	AC0519	UDP-3-O-[3-hydroxy
4	8	1.9	305	2	D85492	UDP-3-O-[3-hydroxy
5	8	1.9	305	2	D90641	UDP-3-O-[3-hydroxy
6	8	1.9	619	2	S67067	probable membrane
7	8	1.9	744	2	B64049	outer membrane pro
8	8	1.9	3488	2	T34418	hypothetical prote
9	7	1.7	50	2	F82656	hypothetical prote
10	7	1.7	68	2	T30399	hypothetical prote
11	7	1.7	89	2	S25601	ubiquinol-cytochro
12	7	1.7	116	2	D97116	uncharacterized co
13	7	1.7	116	2	H95414	hypothetical prote
14	7	1.7	121	2	S17860	phospholipase A2 (
15	7	1.7	126	1	R5DV7	ribosomal protein
16	7	1.7	128	2	D46178	probable transcrip
17	7	1.7	128	2	E46178	probable transcrip
18	7	1.7	138	1	H64434	hypothetical prote
19	7	1.7	151	2	AG1990	hypothetical prote
20	7	1.7	152	2	D72630	hypothetical prote
21	7	1.7	154	2	S76225	hypothetical prote
22	7	1.7	170	2	T25399	hypothetical prote
23	7	1.7	180	2	T28938	hypothetical prote
24	7	1.7	182	2	F83378	hypothetical prote
25	7	1.7	192	2	G71920	hypothetical prote
26	7	1.7	192	2	C64591	conserved hypothet
27	7	1.7	196	2	T15917	hypothetical prote
28	7	1.7	197	2	T18918	hypothetical prote
29	7	1.7	203	2	C87801	protein C10G11.9 [
30	7	1.7	203	2	T25916	hypothetical prote
31	7	1.7	206	2	AB0605	probable tetR-fami
32	7	1.7	217	2	S77510	urease accessory p
33	7	1.7	229	1	W4WL8	E4 protein - human
34	7	1.7	244	2	A72551	hypothetical prote
35	7	1.7	247	2	D69453	hypothetical prote
36	7	1.7	249	2	A83176	hypothetical prote
37	7	1.7	251	2	A24627	calpastatin - pig
38	7	1.7	256	2	AG2324	methionine aminope
39	7	1.7	277	2	JC6142	deoxyguanosine kin
40	7	1.7	287	2	F70361	tRNA-pseudouridine
41	7	1.7	287	2	C75635	phosphoenolpyruvat
42	7	1.7	291	2	T02986	chlorophyll a/b-bi
43	7	1.7	296	2	T40525	hypothetical prote
44	7	1.7	297	1	G69950	6-phosphogluconate
45	7	1.7	297	2	T46414	hypothetical prote

ALIGNMENTS

RESULT 1
A97052
probable lytic murein transglycosylase (N-term. LysM motif repeat domain) CAC1232 [impor
C;Species: Clostridium acetobutylicum
C;Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 09-Jul-2004
C;Accession: A97052
R;Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,
.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A;Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clo
A;Reference number: A96900; MUID:21359325; PMID:21359325
A;Accession: A97052
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-253 <KUR>
A;Cross-references: UNIPROT:Q97JP1; GB:AE001437; PIDN:AAK79204.1; PID:g15024157; GSPDB:G
A;Experimental source: Clostridium acetobutylicum ATCC824
C;Genetics:
A;Gene: CAC1232

Query Match 1.9%; Score 8; DB 2; Length 253;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 108 LLSDLQLA 115
||||||||
Db 94 LLSDLQLA 101

RESULT 2
BVECEA
UDP-3-O-[3-hydroxymyristoyl] N-acetylglucosamine deacetylase (EC 3.5.1.-) - Escherichia
C;Species: Escherichia coli
C;Date: 31-Mar-1988 #sequence_revision 31-Dec-1988 #text_change 09-Jul-2004
C;Accession: A28381; S40606; H64731; C23961
R;Beall, B.; Lutkenhaus, J.
J. Bacteriol. 169, 5408-5415, 1987
A;Title: Sequence analysis, transcriptional organization, and insertional mutagenesis of
A;Reference number: A91852; MUID:88058745; PMID:2824434
A;Accession: A28381
A;Molecule type: DNA
A;Residues: 1-305 <BEA>
A;Cross-references: UNIPROT:P07652; GB:X55034; GB:M10429; NID:g40841; PIDN:CAA38873.1; P
R;Yura, T.; Mori, H.; Nagai, H.; Nagata, T.; Ishihama, A.; Fujita, N.; Isono, K.; Mizobu
submitted to the EMBL Data Library, December 1992
A;Description: Systematic sequencing of the Escherichia coli genome: analysis of the 0-2
A;Reference number: S40531
A;Accession: S40606
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-305 <YUR>
A;Cross-references: EMBL:D10483; NID:g216434; PIDN:BAA01361.1; PID:g216510

R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
.A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A;Title: The complete genome sequence of Escherichia coli K-12.
A;Reference number: A64720; MUID:97426617; PMID:9278503
A;Accession: H64731
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-305 <BLAT>
A;Cross-references: GB:AE000119; GB:U00096; NID:g1786283; PIDN:AAC73207.1; PID:g1786285;
A;Experimental source: strain K-12, substrain MG1655
C;Comment: This protein is involved in cell envelope formation and anomalous cell divisi
.
C;Genetics:
A;Gene: lpxC; envA
A;Map position: 2 min
C;Function:
A;Pathway: lipid A biosynthesis
C;Superfamily: envA protein
C;Keywords: cell division; hydrolase; lipid A biosynthesis

  Query Match 1.9%; Score 8; DB 1; Length 305;
  Best Local Similarity 100.0%; Pred. No. 14;
  Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 18 APFVYLLL 25
||||||||
Db 109 APFVYLLL 116

RESULT 3
AC0519
UDP-3-O-[3-hydroxymyristoyl] N-acetylglucosamine deacetylase [imported] - Salmonella ent
C;Species: Salmonella enterica subsp. enterica serovar Typhi
A;Note: this species has also been called Salmonella typhi
C;Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
C;Accession: AC0519
R;Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,
th, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,
, S.; Moule, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A;Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;
A;Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov
A;Reference number: AB0502; MUID:21534947; PMID:11677608
A;Accession: AC0519
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-305 <PAR>
A;Cross-references: GB:AL513382; PIDN:CAD01291.1; PID:g16501419; GSPDB:GN00176
C;Genetics:
A;Gene: STY0154
C;Superfamily: envA protein

  Query Match 1.9%; Score 8; DB 2; Length 305;
  Best Local Similarity 100.0%; Pred. No. 14;
  Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 18 APFVYLLL 25
||||||||
Db 109 APFVYLLL 116

RESULT 4
D85492
UDP-3-O-[3-hydroxymyristoyl] N-acetylglucosamine deacetylase (EC 3.5.1.-) - Escherichia
C;Species: Escherichia coli
C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 09-Jul-2004
C;Accession: D85492
R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A;Reference number: A85480; MUID:21074935; PMID:11206551
A;Accession: D85492
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-305 <STO>
A;Cross-references: UNIPROT:P07652; GB:AE005174; NID:g12512802; PIDN:AAG54400.1; GSPDB:G
A;Experimental source: strain O157:H7, substrain EDL933
C;Genetics:
A;Gene: lpxC
C;Superfamily: envA protein
C;Keywords: hydrolase

  Query Match 1.9%; Score 8; DB 2; Length 305;
  Best Local Similarity 100.0%; Pred. No. 14;
  Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 18 APFVYLLL 25
||||||||
Db 109 APFVYLLL 116

RESULT 5
D90641
UDP-3-O-[3-hydroxymyristoyl] N-acetylglucosamine deacetylase (EC 3.5.1.-) - Escherichia
C;Species: Escherichia coli
C;Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 09-Jul-2004
C;Accession: D90641
R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and geno
A;Reference number: A99629; MUID:21156231; PMID:11258796
A;Accession: D90641
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-305 <HAY>
A;Cross-references: UNIPROT:P07652; GB:BA000007; PIDN:BAB33523.1; PID:g13359556; GSPDB:G
A;Experimental source: strain O157:H7, substrain RIMD 0509952
C;Genetics:
A;Gene: ECs0100
C;Superfamily: envA protein
C;Keywords: hydrolase

  Query Match 1.9%; Score 8; DB 2; Length 305;
  Best Local Similarity 100.0%; Pred. No. 14;
  Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 18 APFVYLLL 25
||||||||
Db 109 APFVYLLL 116

RESULT 6
S67067
probable membrane protein YOR175c - yeast (Saccharomyces cerevisiae)
N;Alternate names: hypothetical protein O3635
C;Species: Saccharomyces cerevisiae
C;Date: 12-Jul-1996 #sequence_revision 12-Jul-1996 #text_change 09-Jul-2004
C;Accession: S67067; S67063
R;Hughes, B.; Pohl, T.M.
submitted to the Protein Sequence Database, July 1996
A;Reference number: S66685
A;Accession: S67067
A;Molecule type: DNA
A;Residues: 1-619 <HUG>
A;Cross-references: UNIPROT:Q08548; EMBL:Z75083; NID:g1420424; PID:e252056; PID:g1420425
A;Experimental source: strain S288C
R;Bordonne, R.; Camasses, A.; Madania, A.; Martin, R.P.; Poch, O.; Tarassov, I.A.; Winso
submitted to the Protein Sequence Database, July 1996
A;Reference number: S67032
A;Accession: S67063
A;Molecule type: DNA
A;Residues: 270-619 <BOR>
A;Cross-references: EMBL:Z75083; MIPS:YOR175c

A;Experimental source: strain S288C
C;Genetics:
A;Cross-references: SGD:S0005701
A;Map position: 15R
C;Keywords: transmembrane protein
F;53-69/Domain: transmembrane #status predicted <TM1>
F;461-477/Domain: transmembrane #status predicted <TM2>

Query Match 1.9%; Score 8; DB 2; Length 619;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 268 RKQKKLEK 275
||||||||
Db 500 RKQKKLEK 507

RESULT 7
B64049
outer membrane protein hxuC homolog - Haemophilus influenzae (strain Rd KW20)
C;Species: Haemophilus influenzae
C;Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 09-Jul-2004
C;Accession: B64049
R;Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A
; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, J
, D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.M.
Science 269, 496-512, 1995
A;Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter,
A;Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
A;Reference number: A64000; MUID:95350630; PMID:7542800
A;Accession: B64049
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-744 <TIGR>
A;Cross-references: UNIPROT:P44523; GB:U32696; GB:L42023; NID:g1573057; PIDN:AAC21789.1;

Query Match 1.9%; Score 8; DB 2; Length 744;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 250 IFLLVTLV 257
||||||||
Db 9 IFLLVTLV 16

RESULT 8
T34418
hypothetical protein F12F3.3 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C;Accession: T34418
R;Fulton, B.; Wohldmann, P.
submitted to the EMBL Data Library, July 1998
A;Description: The sequence of C. elegans cosmid F12F3.
A;Reference number: Z21521
A;Accession: T34418
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-3488 <FUL>
A;Cross-references: EMBL:U80022; PIDN:AAC25885.1; GSPDB:GN00023; CESP:F12F3.3
A;Experimental source: strain Bristol N2; clone F12F3
C;Genetics:
A;Gene: CESP:F12F3.3
A;Map position: 5
A;Introns: 281/3; 332/1; 562/3; 600/3; 1866/3; 1944/3; 3393/1

Query Match 1.9%; Score 8; DB 2; Length 3488;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 180 WLKDGKPL 187
||||||||
Db 1797 WLKDGKPL 1804

RESULT 9
F82656
hypothetical protein XF1631 [imported] - Xylella fastidiosa (strain 9a5c)
C;Species: Xylella fastidiosa
C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 09-Jul-2004
C;Accession: F82656
R;anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequer
Nature 406, 151-157, 2000
A;Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A;Reference number: A82515; MUID:20365717; PMID:10910347
A;Note: for a complete list of authors see reference number A59328 below
A;Accession: F82656
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-50 <SIM>
A;Cross-references: UNIPROT:Q9PCX3; GB:AE003990; GB:AE003849; NID:g9106683; PIDN:AAF8444
A;Experimental source: strain 9a5c
R;Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A
Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer, H
as-Neto, E.; Docena, C.; El-Dorry, H.; Facincani, A.P.; Ferreira, A.J.S.
submitted to GenBank, June 2000
A;Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laigr
chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E
A;Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.,
, F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak
A;Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir
M.; Tsuhako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z
A;Reference number: A59328
A;Contents: annotation
C;Genetics:
A;Gene: XF1631

Query Match 1.7%; Score 7; DB 2; Length 50;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 189 NDSRMLL 195
|||||||
Db 28 NDSRMLL 34

RESULT 10
T30399
hypothetical protein ORF52 - Lymantria dispar nuclear polyhedrosis virus
C;Species: Lymantria dispar nuclear polyhedrosis virus, LdMNPV
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 09-Jul-2004
C;Accession: T30399
R;Kuzio, J.; Pearson, M.N.; Harwood, S.H.; Funk, C.J.; Evans, J.T.; Slavicek, J.M.; Rohr
Virology 253, 17-34, 1999
A;Title: Sequence and analysis of the genome of a baculovirus pathogenic for Lymantria d
A;Reference number: Z20836; MUID:99124785; PMID:9887315
A;Accession: T30399
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-68 <KUZ>
A;Cross-references: UNIPROT:Q9YMS2; EMBL:AF081810; PIDN:AAC70237.1

Query Match 1.7%; Score 7; DB 2; Length 68;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 363 RSPARSP 369
|||||||
Db 32 RSPARSP 38

RESULT 11
S25601
ubiquinol-cytochrome-c reductase (EC 1.10.2.2) cytochrome b - brine shrimp mitochondrion

C;Species: mitochondrion Artemia sp. (brine shrimp)
A;Variety: strain Delta del Ebro
C;Date: 22-Nov-1993 #sequence_revision 19-Oct-1995 #text_change 03-Jun-2002
C;Accession: S60630; S25601
R;Perez, M.L.; Valverde, J.R.; Batuecas, B.; Amat, F.; Marco, R.; Garesse, R.
J. Mol. Evol. 38, 156-168, 1994
A;Title: Speciation in the Artemia genus: mitochondrial DNA analysis of bisexual and par
A;Reference number: S60624; MUID:94223692; PMID:8169960
A;Accession: S60630
A;Molecule type: DNA
A;Residues: 1-89 <PE2>
A;Cross-references: EMBL:X67260; NID:g11220; PIDN:CAA47682.1; PID:g11221
A;Experimental source: strain Delta del Ebro
A;Note: the source is designated as Artemia parthenogenetica
C;Genetics:
A;Gene: cytb
A;Genome: mitochondrion
A;Genetic code: SGC4
C;Superfamily: cytochrome b; cytochrome b homology; cytochrome b6 homology; plastoquinol
C;Keywords: chromoprotein; electron transfer; heme; iron; metalloprotein; mitochondrion;
F;1-89/Domain: cytochrome b homology (fragment) <CBH>
F;1-28/Domain: cytochrome b6 homology (fragment) <CB6>
F;1-19/Domain: transmembrane (fragment) #status predicted <TM4>
F;40-89/Domain: plastoquinol-plastocyanin reductase 17K protein homology (fragment) <17K
F;48-64/Domain: transmembrane #status predicted <TM5>
F;1/Binding site: heme iron (His) (axial ligand) (low potential) #status predicted
F;15/Binding site: heme iron (His) (axial ligand) (high potential) #status predicted

  Query Match 1.7%; Score 7; DB 2; Length 89;
  Best Local Similarity 100.0%; Pred. No. 45;
  Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 250 IPLLVTL 256
|||||||
Db 55 IPLLVTL 61

RESULT 12
D97116
uncharacterized conserved protein, YLXM B. subtilis ortholog CAC1753 [imported] - Clostr
C;Species: Clostridium acetobutylicum
C;Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 09-Jul-2004
C;Accession: D97116
R;Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,
.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A;Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clo
A;Reference number: A96900; MUID:21359325; PMID:21359325
A;Accession: D97116
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-116 <KUR>
A;Cross-references: UNIPROT:Q97I99; GB:AE001437; PIDN:AAK79719.1; PID:g15024722; GSPDB:G
A;Experimental source: Clostridium acetobutylicum ATCC824
C;Genetics:
A;Gene: CAC1753
C;Superfamily: hypothetical protein A05_orf102

  Query Match 1.7%; Score 7; DB 2; Length 116;
  Best Local Similarity 100.0%; Pred. No. 57;
  Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 310 LYILKDK 316
|||||||
Db 88 LYILKDK 94

RESULT 13
H95414
hypothetical protein SMa2273 [imported] - Sinorhizobium meliloti (strain 1021) magaplasm
C;Species: Sinorhizobium meliloti
C;Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 09-Jul-2004
C;Accession: H95414
R;Barnett, M.J.; Fisher, R.F.; Jones, T.; Komp, C.; Abola, A.P.; Barloy-Hubler, F.; Bows
.; Kalman, S.; Keating, D.H.; Palm, C.; Peck, M.C.; Surzycki, R.; Wells, D.H.; Yeh, K.C.
Proc. Natl. Acad. Sci. U.S.A. 98, 9883-9888, 2001
A;Title: Nucleotide sequence and predicted functions of the entire Sinorhizobium melilot
A;Reference number: A95262; MUID:21396509; PMID:11481432
A;Accession: H95414
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-116 <KUR>
A;Cross-references: UNIPROT:Q92XM0; GB:AE006469; PIDN:AAK65882.1; PID:g14524391; GSPDB:G
A;Experimental source: strain 1021, megaplasmid pSymA
R;Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler,
pela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;
L.; Hyman, R.W.; Jones, T.
Science 293, 668-672, 2001
A;Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure,
hebault, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.
A;Title: The composite genome of the legume symbiont Sinorhizobium meliloti.
A;Reference number: A96039; MUID:21368234; PMID:11474104
A;Contents: annotation
C;Genetics:
A;Gene: SMa2273
A;Genome: plasmid

  Query Match 1.7%; Score 7; DB 2; Length 116;
  Best Local Similarity 100.0%; Pred. No. 57;
  Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 88 IGTLRPD 94
|||||||
Db 61 IGTLRPD 67

RESULT 14
S17860
phospholipase A2 (EC 3.1.1.4) isoform A1 - leaf-nosed viper
C;Species: Eristocophis macmahoni (leaf-nosed viper)
C;Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 09-Jul-2004
C;Accession: S17860
R;Siddiqi, A.R.; Zaidi, Z.H.; Joernvall, H.
Eur. J. Biochem. 201, 675-679, 1991
A;Title: Purification and characterization of two highly different group II phospholipas
A;Reference number: S17860; MUID:92037623; PMID:1935962
A;Accession: S17860
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1-121 <SID>
A;Cross-references: UNIPROT:P24293
C;Superfamily: phospholipase A2
C;Keywords: carboxylic ester hydrolase

  Query Match 1.7%; Score 7; DB 2; Length 121;
  Best Local Similarity 100.0%; Pred. No. 59;
  Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 48 GKSALLS 54
|||||||
Db 14 GKSALLS 20

RESULT 15
R5DV7
ribosomal protein L7/L12 - Desulfovibrio vulgaris (strain Miyazaki)
C;Species: Desulfovibrio vulgaris
C;Date: 17-May-1985 #sequence_revision 17-May-1985 #text_change 09-Jul-2004
C;Accession: A02769
R;Itoh, T.; Otaka, E.
Biochim. Biophys. Acta 789, 229-233, 1984
A;Title: Complete amino-acid sequence of an L7/L12-type ribosomal protein from Desulfovi
A;Reference number: A02769
A;Accession: A02769
A;Molecule type: protein
A;Residues: 1-126 <ITO>

A;Cross-references: UNIPROT:P02393
C;Superfamily: Escherichia coli ribosomal protein L12
C;Keywords: methylated amino acid; protein biosynthesis; ribosome
F;76,87/Modified site: N6-methyllysine (Lys) #status experimental

Query Match 1.7%; Score 7; DB 1; Length 126;
Best Local Similarity 100.0%; Pred. No. 61;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 156 TVLELSE 162
|||||||
Db 16 TVLELSE 22

Search completed: August 3, 2005, 10:10:07
Job time : 43 secs